C3.261:80-3

SPECIAL
DEMOGRAPHIC
ANALYSES

CDS-80-3

Measurement of Subjective Phenomena

U. S. Department of Commerce
BUREAU OF THE CENSUS

SPECIAL
DEMOGRAPHIC
ANALYSES

CDS-80-3

Measurement of Subjective Phenomena

Denis F. Johnston, Editor

Issued October 1981

U.S. Department of Commerce
Malcolm Baldrige, Secretary
Joseph R. Wright, Jr., Deputy Secretary
Robert G. Dederick, Assistant Secretary
for Economic Affairs

BUREAU OF THE CENSUS
Daniel B. Levine, Acting Director

BUREAU OF THE CENSUS
Daniel B. Levine, Acting Director

CENTER FOR DEMOGRAPHIC STUDIES
James R. Wetzel, Chief

Library of Congress Cataloging in Publication Data

Main entry under title:

Measurement of subjective phenomena.

(Special demographic analyses; CDS-80-3)
"Issued October 1981."
Bibliography: p.
Contents: Introduction/Denis F. Johnston—Dissatisfaction with satisfaction: subjective social indicators and the quality of life/Allen R. Wilcox—Comments/Frank M. Andrews—[etc.]
1. Social indicators—United States—Addresses, essays, lectures. 2. Quality of life—United States—Addresses, essays, lectures. 3.Social surveys—United States—Addresses, essays, lectures. I. Johnston, Denis Foster. II. Series.
HN60.M42 301'.0723 81-12247
AACR2

For Sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402

Preface

This monograph continues the series, Special Demographic Analyses, issued by the Center for Demographic Studies, Bureau of the Census. The purpose of these publications is to provide insight and perspective on important demographic trends and patterns. However, that insight cannot be obtained by considering conventional demographic topics alone. Demographic processes, like all social processes, are associated with and affected by a complex variety of socioeconomic and cultural conditions, many of which exert their influence indirectly, as mediated by human values, attitudes, perceptions, and subjective reactions or feelings. Therein lies the relevance of the topic addressed by the several essays in this monograph. The primary concern of these essays is the methodological issues that have emerged in our attempts to ascertain and understand the bewildering variety of attitudinal factors that mediate many aspects of human life and behavior.

The monograph includes four major essays by Allen R. Wilcox, Charles F. Turner, Donald C. Dahmann, and Tom W. Smith. Because of their critical orientation, the first two essays (Wilcox and Turner) are supplemented by commentaries by Frank M. Andrews and Angus Campbell, respectively. Replies by the original authors are also included. The monograph concludes with a brief afterword by Edwin D. Goldfield.

The contributors to this monograph represent a rich diversity of experience in the measurement of subjective phenomena and a wide range of professional interests in this challenging field of inquiry. Allen R. Wilcox, the author of the first essay, is professor of political science and director of the Bureau of Governmental Research at the University of Nevada, Reno. He is the coauthor of *Legislative Roll-Call Analysis* (Northwestern University Press, 1966) and the editor and coauthor of *Public Opinion and Political Attitudes* (John Wiley & Sons, 1974). He is also general editor of the *Nevada Public Affairs Review* and has published articles in the *American Behavioral Scientist, Western Political Quarterly, Policy Studies Journal,* and *Behavioral Science,* among others. His current research projects include a statistical monograph, a book linking personality theory to

political theory and related ideas in other disciplines, and survey research on urban political processes.

The commentary on the Wilcox essay is by Frank M. Andrews, research scientist and program director, Survey Research Center, Institute for Social Research, and professor of psychology and population planning, University of Michigan. His many contributions to the field of attitude measurement include two major studies: *Social Indicators of Well-Being: Americans' Perceptions of Life Quality* (coauthored with Stephen B. Withey, Plenum Press, 1976) and *Quality of Life—Comparative Studies* (coedited with Alexander Szalai, Sage, 1980).

Charles F. Turner is a relative newcomer to the field, having completed his dissertation on the social psychology of socioeconomic attainment at Columbia University in 1978. He is currently senior staff officer and study director for the Committee on National Statistics of the National Research Council, National Academy of Sciences, where he is directing the Committee's study of survey measurements of subjective phenomena under the guidance of a panel of experts chaired by Otis Dudley Duncan.

The commentary on Turner's essay is by the late Angus Campbell, whose reputation in the field of public opinion studies and quality-of-life studies is worldwide. He was director of the Survey Research Center, Institute for Social Research, University of Michigan, from 1970 to 1976 and continued to serve there as a program director. His major works include *The Human Meaning of Social Change* (coedited with Philip E. Converse, Russell Sage Foundation, 1972), *The Quality of American Life: Perceptions, Evaluations and Satisfactions* (coauthored with Philip E. Converse and Willard L. Rodgers, Russell Sage Foundation, 1976), and *The Sense of Well-Being in America: Recent Patterns and Trends* (McGraw-Hill, 1981). The most recent of the many awards and honors he received during his distinguished career was his election to membership in the National Academy of Sciences.

The third essay is by Donald C. Dahmann, a staff member of the Center for Demographic Studies, Bureau of the Census. After receiving his doctorate in geography from the University of Chicago in 1976, he spent a year on the staff of the National Research Council, National Academy of Sciences. He then joined the Bureau of the Census. His recent publications are "Population Redistribution in the United States in the 1970's" (coauthored with Brian J. L. Berry in Berry and Silverman, *Population Redistribution and Public Policy*, National Academy of Sciences, 1980) and *The City-Suburb Income Gap: Is It Being Narrowed by a Back-to-the-City Movement?* (coauthored with Larry H. Long and first in this series, March 1980).

The final essay is by Tom W. Smith, senior study director, National Opinion Research Center, University of Chicago. Trained as an historian, Dr. Smith has been especially interested in the study of social change and in survey research methodology. He has performed trend studies of such diverse topics as feminism, happiness, and crime. He has also carried out a number of studies of factors affecting survey results and related methodological issues.

Edwin D. Goldfield, author of the afterword, is executive director of the Committee on National Statistics, National Research Council, National Academy of Sciences. The editor, Denis F. Johnston, is senior advisor on social indicators, Center for Demographic Studies, Bureau of the Census.

Digitized by the Internet Archive
in 2012 with funding from
LYRASIS Members and Sloan Foundation

http://archive.org/details/measurementofsub00john

Contents

Introduction ---------- ix
Denis F. Johnston

Dissatisfaction With Satisfaction: Subjective Social Indicators and the Quality of Life ---------- 1
Allen R. Wilcox

Comments ---------- 21
Frank M. Andrews

Response to Comments ---------- 31
Allen R. Wilcox

Surveys of Subjective Phenomena: A Working Paper ---------- 37
Charles F. Turner

Irregularities in Survey Data ---------- 79
Angus Campbell

Patterns of Disagreement: A Reply to Angus Campbell ---------- 85
Charles F. Turner

Subjective Indicators of Neighborhood Quality ---------- 97
Donald C. Dahmann

Can We Have Confidence in Confidence? Revisited ----------119
Tom W. Smith

Afterword ---------- 191
Edwin D. Goldfield

Introduction

The Bureau of the Census's interest in measures of what are termed here *subjective phenomena* is entirely practical. In the first place, the Bureau is necessarily concerned with maintaining and improving the quality of public responses to its questionnaires and other data-gathering procedures. That aim entails a continuing effort to understand the variety of attitudinal and emotional factors that can impede effective communication between the respondent and the information-seeker.

In addition, the social indicators research and development program at the Bureau's Center for Demographic Studies has generated renewed interest in the relationship between objective conditions and characteristics of the population and public perceptions of those conditions. In preparing its triennial social indicators reports, the Center has sought to present statistical information on both the objective and subjective aspects of the principal components of daily life, such as health, education, housing, work, income, and the like. Hence the Bureau is vitally interested in the development of criteria whereby the more valid and reliable of these subjective measures could be differentiated.

About 10 years ago, Angus Campbell and Philip E. Converse (1972) tried to explain the persistent demand for measures of subjective phenomena. Their fundamental argument was to the effect that the social significance of many objective conditions could not be ascertained without taking into account their perceived meaning and importance to the people directly involved. They pointed out, in addition, that many social processes and changes are strongly influenced by changing values and aspirations, which are essentially subjective in nature. Without denying the reciprocal influence of objective circumstances upon these values, they asserted the continuing role of human aspirations in motivating sustained efforts toward the achievement of specified goals. They also called attention to the fact that some of the more intriguing and promising theoretical speculations designed to explain important aspects of human behavior postulated a variety of sociopsychological states or propensities whose existence and effects demanded some kind of empirical verification. They buttressed their argument for

the relevance of subjective phenomena by stressing the problem of choice, especially as it has emerged in modern affluent societies. By this view, concomitant increases in education, material affluence, leisure, and the manifold opportunities afforded by urban life have enabled many individuals to experience a range of choice that was unknown to all but the most wealthy and powerful individuals of a few decades past. But if this freedom of choice has liberated us from the traditional determinisms of location, rank, and material circumstances, it has also posed a threat to traditional norms and values. As Erich Fromm's now-classic *Escape From Freedom* made clear, the opportunity to exercise one's personal preferences over an unprecedented variety of objects, ranging from trivial matters of dress or diet to fundamental value choices regarding occupational careers and lifestyles, has been a mixed blessing for many. But whatever its costs and benefits, it is the existence of choice that underlies the importance of personal feelings, attitudes, reactions, and aspirations.

Finally, they emphasized the policy relevance of subjective states and processes, particularly where public attitudes and responses play a critical role in determining policy outcomes. Their argument here was expressed a half-century earlier in W. I. Thomas's dictum that it is not the objective situation that determines the nature of human responses but rather how that situation is perceived by the actors in question. Policymakers intuitively recognize this point, hence their persistent concern with public views and reactions.

These arguments are both familiar and persuasive. But if they tell us that public perceptions, interpretations, and reactions are among the things that ought to be taken into account in any realistic assessment of human conditions, they do not tell us how or even whether such measurement can be carried out. The large and growing body of literature relating to opinion polling and related methodological issues offers ample evidence that our ability to ascertain these public feelings and attitudes with reasonable precision and reliability leaves much to be desired. The conviction that public perceptions, feelings, and reactions are the right things to measure is of small comfort in the face of evidence that such measures are unreliable or impossible to interpret. In short, the prevailing situation in this challenging field is decidedly unsatisfactory except to those who regard mental states as epiphenomenal or illusory. If we believe, on the contrary, that subjective states and feelings are significant independent factors in shaping human behavior, we are bound to regard the measurement of subjective phenomena as an important objective of social science research.

The contributors to this monograph obviously subscribe to the view that subjective phenomena are important objects of scientific analysis. In addition, some of them would probably agree with the proposition that the search for improved measures of subjective phenomena has significant implications for the improvement of our objective measures as well. The underlying argument here is that the distinction between objective and subjective conditions or phenomena is much sharper in theory than in reality. Simplified accounts of measurement procedures applied to objective phenomena describe the process as one of

applying an agreed criterion measure (e.g., a tape measure) to determine a particular dimension (e.g., length) of a given object (e.g., a bookcase). A more illuminating example, however, recognizes the play of subjective elements at several critical stages between the occurrence of some stimulus and the establishment of some datum relating thereto. As expressed by Johan Galtung, the sequence from stimulus to datum proceeds in five stages along two converging paths. The first path is an "outside world" sequence: stimulus to object to response to impression to datum. The second is a corresponding sequence followed by the observer: presentation (of the stimulus) to manifestation (as a coherent entity) to perception to recording (of impressions) to datum (Galtung, 1967, p. 27). To this account, Neville Layne adds the critical observation that perception is not limited to what is real (i.e., to what exists) but also reflects what is important. Layne (1977, p. 6) goes on to distinguish between technical perception, concerned with the nature of the real, and theoretical perception, focused on questions of importance or relevance.

These distinctions may help us to understand how the interests of researchers and those of policymakers can be reconciled. The former seek to deepen our understanding of the real world, its characteristics, dimensions, and so on. They are therefore primarily concerned with precise measurement in their quest for reliable generalizations. But policymakers are more sensitive to significance (in the nonstatistical sense) than to precision; they wish to know what counts, not what is countable. However, it is becoming evident to many researchers and policymakers alike that these interests are complementary. Most researchers are interested in seeing their results find useful application in the real world. Hence their growing sensitivity to the practical concerns of policymakers. For their part, policymakers are increasingly aware that valid understanding is a requisite to rational decisionmaking.

The present state of the art is a matter of some controversy. The general impression one obtains from the essays that follow is that our measures of subjective phenomena are often unable to represent adequately the attitudes or perspectives they purport to reflect. But if that is indeed the case, it might be well to bear in mind that many of our most familiar objective measures also need improvement and that such improvement entails, in many cases, some infusion of subjective elements into the measurement process. For example, it is certainly possible to ascertain the educational attainment of any population in terms of the familiar "years of school completed." Such data are routinely collected at minimal cost and with a high degree of reliability. But one can certainly ask whether such data are sufficient to reflect educational attainment. More precisely, one can question the adequacy of the conception of educational attainment that is reflected in such data. If we agree that educational attainment involves far more than the duration of one's formal schooling, that it encompasses such considerations as one's functional adequacy or ability to cope with life's exigencies, we are but a small step from recognizing the need to include a host of subjective factors, such as one's self-perceptions, feelings of efficacy, self-confidence, and the like.

The difficulty with many simple, inexpensive, familiar, and objective measures is that rigorous operational definitions of the phenomena they reflect typically fall far short of the complex reality that is of interest to researchers and policymakers alike. Hence the familiar plight: reliance upon familiar measures of limited validity versus the time-consuming and expensive development of more elaborate constructs whose reliability is questionable.

Peter Winch (1958) has described the nature of historical explanation in terms that capture the essence of the dissatisfaction with purely objective measures among those who seek to assess the conditions of human life:

> Historical explanation is not the application of generalizations and theories to particular instances: it is the tracing of internal relations. It is like applying one's knowledge of a language in order to understand a conversation rather than like applying one's knowledge of the laws of mechanics to understand the workings of a watch. (p. 133)

Angus Campbell was never willing to surrender the study of human behavior to either the case study methods of the historians or the mechanical quantifications of the model builders. His aim was to combine the sensitivity of the one with the explanatory power of the other. His efforts stand as an example and a challenge to those who would pursue the same goal; hence this monograph is respectfully dedicated to his memory, in the hope that it will contribute to and stimulate further development in the field.

Denis F. Johnston
Bureau of the Census

REFERENCES

Campbell, Angus, and Philip E. Converse. 1972. "Social Change and Human Change." In Angus Campbell and Philip E. Converse, eds., *The Human Meaning of Social Change.* New York: Russell Sage Foundation, pp. 1-16.

Galtung, Johan. 1967. *Theory and Methods of Social Research.* Oslo: Universitetsforlaget.

Layne, Neville. March 1977. *Perceptual Indicators: A Conceptualization.* Ottawa: Policy Research and Long Range Planning Branch (Welfare), Department of National Health and Welfare, Staff Working Paper 7703, p. 6.

Winch, Peter. 1958. *The Idea of a Social Science and Its Relation to Philosophy.* London: Routledge & Kegan Paul.

Dissatisfaction With Satisfaction: Subjective Social Indicators and the Quality of Life

Allen R. Wilcox
University of Nevada, Reno

INTRODUCTION

Early in the development of the social indicators movement in the United States and elsewhere, researchers began to distinguish between objective and subjective indicators. The relationship (or lack thereof) between the two types has been explored in a number of studies (e.g., Stipak, 1979; Allardt, 1978; Schneider, 1976). The consensus that has emerged is that they are both useful and needed and should somehow be employed to complement each other.

One difficulty with this consensus is that the distinction between objective and subjective indicators has not been clearly drawn (Andrews and Withey, 1976, pp. 5-6), so that the nature of the consensus is also unclear. In this chapter, subjective social indicators are defined crudely and operationally as the responses of target populations to questions about their beliefs, attitudes, values, and other internal (subjective) states. One interpretation of such subjective responses has been that they are measuring the respondent's quality of life and, through aggregation, the quality of life of the society of which they are a part. These interpretations and others are subjected to critical scrutiny in this chapter.

The specific focus of the chapter is two major studies of subjective social indicators conducted by different cadres of researchers at the University of Michigan—*The Quality of American Life: Perceptions, Evaluations, Satisfactions* by Campbell, Converse, and Rodgers (1976) and *Social Indicators of Well-Being: Americans' Perceptions of Life Quality* by Andrews and Withey (1976).

Both of these studies, although clearly pathbreaking and sophisticated in design, analysis, and interpretation, rely heavily on questions regarding how satisfied respondents are with their lives and various aspects of their lives. It is this

This is a revised version of a paper presented at the 74th Annual Meeting of the American Sociological Association, Boston, August 27-31, 1978. I wish to thank Tom Atkinson and Frank M. Andrews for sharing with me, in the best scholarly tradition, their latest work on subjective social indicators.

reliance on questions about satisfaction[1] with which we are dissatisfied. Because this field of research is one of the most important currently being mined, we believe it is necessary to raise all possible questions about it that can lead to cumulative improvement. The comments we make below are grouped into three levels of ascending generality roughly responsive to these questions: Do these measures of satisfaction adequately measure satisfaction? Do they adequately capture respondents' *perceived* quality of life? And, assuming favorable responses to the first two questions, What sense do they provide of quality of life per se?

LEVEL 1: SATISFACTION

The primary measures used in the two studies are different. In *The Quality of American Life* (QOAL), respondents were asked how satisfied they were with various aspects of life. There were seven response categories, with only three labeled (1 = completely satisfied, 4 = neutral, 7 = completely dissatisfied). In *Social Indicators of Well-Being* (SIWB), respondents were asked how they felt about various aspects of life. The response categories were more complex. Responses could fall on a 7-point scale that had all categories labeled (1 = terrible, 2 = unhappy, 3 = mostly dissatisfied, 4 = mixed—about equally satisfied and dissatisfied, 5 = mostly satisfied, 6 = mostly pleased, 7 = delighted) or into any one of three off-scale categories (A = neutral—neither satisfied nor dissatisfied, B = I never thought about it, C = does not apply to me). The authors of SIWB clearly gave extensive thought to their format and among others reasons chose it to avoid the markedly skewed distributions often obtained in the QOAL study. However, it is this delighted-terrible (D-T) scale, in particular, that raises several questions.

The Delighted and the Terrible

From a semantic perspective, the labels given to the seven categories of the D-T scale are quite problematic:

1. The antonyms for pleased and mostly satisfied are privative opposites (i.e., opposites denoting the absence of a property), but the antonym for delighted is not, rendering the antonymic construction inconsistent.[2]
2. It is not obvious that pleased is semantically more extreme than mostly satisfied or that unhappy is more extreme than mostly dissatisfied.
3. Pleased and unhappy are weak antonyms at best. Why displeased or happy were not used is not discussed. (If happy were used, is there any difference in using sad as opposed to unhappy as an antonym?)

---

[1] Not all questions we are concerned with literally ask about *satisfaction*. To avoid unduly cumbersome prose, we use satisfaction on occasion as a terminological stand-in for a cluster of similar types of questions. The context should make the usage clear.

[2] The terminology is from Lyons's *Semantics* (1977, p. 279). To investigate the problems raised, formal semantic theory must be supplemented by psycholinguistic studies such as those on the intensity of adjectives summarized in Clark and Clark (1977, pp. 458-459).

4. Delighted and terrible are questionable antonyms. A lengthy journey through *Roget's Thesaurus* turned up an extremely indirect link. Terrible, in particular, appears to be a distant semantic cousin of all the other labels.
5. The D-T scale as a whole incorporates the ideas of delight, terribleness, satisfaction, pleasure, and happiness, all in one supposedly unidimensional scale. This appears to stretch the concept of unidimensionality beyond recognition.[3]

Labeling

The authors of SIWB defend the use of explicit labels for all categories of their scale (Andrews and Withey, 1979, p. 20) on the grounds that it lessens ambiguity. Explicit labels, of course, are hardly unique to the D-T scale, yet the procedure may need rethinking. Recent work by Bradburn and Miles (Bradburn and Sudman, 1979, pp. 152-162) suggests that the use of imprecise quantifying words creates a number of problems and that the substitution of numerical estimates increases the explanatory power of the items in question.

Category Scaling

In a recent article, Lodge and Tursky (1979) examine the common use of category scaling—where the respondent rates an item or expresses a judgment by selecting one of a fixed number of options—and suggests the procedure has a number of serious weaknesses. Specifically, (1) "information is lost through the limited resolution of the categories," (2) "category scales represent only an ordinal level of measurement, thereby limiting the statistical procedures which can be used without reservation," and (3) "the investigator is inadvertently affecting the response—artifactually constraining or expanding the range which can be used by the respondent" (Lodge and Tursky, 1979, p. 50). (The authors also comment on labeling.) This last weakness is particularly relevant to QOAL and SIWB because of the authors' difficulty in obtaining significant variance in their frequency distributions. Relying on years of research in psychophysics, Lodge and Tursky advocate and attempt to demonstrate the use of magnitude scaling in surmounting, or at least mitigating, these weaknesses. Their work should be consulted for the details.

Positivity Bias

Positivity bias is the phrase used in QOAL to refer to the tendency of respondents to report greater happiness, satisfaction, well-being, and so on than they truly feel. In an article on happiness, Tom W. Smith divides this tendency

[3] A recent cross-national study of subjective well-being by Andrews and Inglehart (1979) finds differences between the United States and Western European countries. However, the U.S. data are derived from the D-T scale, while the European data are derived from an 11-point satisfaction scale. Given the set of comments in the text, one might wonder to what extent the observed differences are a semantic artifact.

into several subphenomena: (1) the basic positive orientation of American culture, (2) the consistent rating of self-happiness higher than happiness of others, and (3) social desirability-conformance with a social norm dictating expressions of satisfaction or happiness, especially evident in personal interview situations (Smith, 1979, p. 19).[4]

Such a bias, whatever its derivation, has serious implications. It may lead to overestimation of satisfaction and other positive states in the general population, misjudgment of comparative satisfaction among subpopulations, and misinterpretation of changes in satisfaction in both general and subpopulations. The authors of QOAL and SIWB are quite aware of these possibilities, and QOAL, in particular, includes an entire chapter agonizing over the problem. There is no need to recapitulate that discussion and analysis. The judgment here is that no satisfactory way of controlling for or estimating positivity bias has yet been demonstrated.

Self-knowledge

The social and cultural flavor of the discussions of positivity bias shade over into the personality-oriented comments treated here under the rubric of *self-knowledge*. The psychoanalytic concept of *repression* has been but one approach to the phenomenon of being in some sense unaware of one's true feelings. The concept of *self-deception*, recently investigated in the experimental literature, is another (Gur and Sackheim, 1979). Anyone reading a report such as Grier and Cobb's *Black Rage* (1968) is unlikely to doubt the possibility of surface satisfaction covering undercurrents of hate, resentment, and other emotions unlikely to be elicited in the standard survey interview (see also Solomon, 1976, chapter 13).

In large part we are referring here to the difficulties of obtaining valid information from sample surveys and particularly from forced-choice questions. One set of alternatives implied by the preceding comments is the array of interpretative devices employed by clinical psychologists. Within sociology, the weaknesses of surveys have been examined repeatedly, with a particularly useful discussion to be found in the concise treatment by Derek Phillips in *Knowledge from What?* (1971). The seminal corpus of work dealing with alternatives derives from Donald T. Campbell and his associates, work reorganized under the heading of "triangulation" by Norman K. Denzin (1970) and H. W. Smith (1975) (see also Leege and Francis, 1974, p. 140).

In the present context, the strategy suggested by these methodologists would mean zeroing in on true satisfaction by capitalizing on the unique strengths of alternative methodologies. The authors of QOAL and SIWB do employ it at the level of alternative ways of eliciting forced-choice responses. To our knowledge,

[4] See the speculations on goodness as a normal state in Clark and Clark (1977, p. 539). The earlier literature on social desirability is obtainable through Sudman and Bradburn (1974).

a more far-reaching attempt at triangulation in an effort to get past the problems of positivity bias and lack of self-knowledge has yet to be initiated.

Models of Evaluation

The authors of QOAL and SIWB are well aware that feelings of satisfaction may be the result of a number of different psychological processes. In fact, SIWB devotes a chapter to the topic, enumerating a series of "models for evaluation" and presenting the results of several exploratory analyses (pp. 219-220) (see also Campbell et al., 1976). The models include orientations toward recent improvement, ideal standards, neutral or "swing" points, pace of change, and so on. Earlier in the text, the authors mention several bodies of literature relevant to these models--external-internal interaction, adaption-level, social judgment, and utility functions (pp. 14-18).

One body of literature not explicity mentioned (though probably subsumable under the social judgment category) is that relating to the concept of relative deprivation, which itself has been embedded in several different formulations (Crosby, 1979). To develop the present argument, we shall dip briefly into a contribution to this literature by Robin Williams and extract one summary quotation:

1. To receive less than one *wants* (desires, needs) results in a sense of *deprivation*
2. To receive less than one *expects* results in feeling *disappointment*
3. To receive less than is *mandated* by accepted social rules and values (that to which one is entitled) results in a sense of *injustice* (1975, p. 356)

We assume the following. The same level of dissatisfaction might be registered by respondents, some of whom are responding in terms of deprivation, some in terms of disappointment, and some in terms of injustice. We further assume that the consequences and implications of these different cases are not identical. Injustice-dissatisfaction may be more likely to lead to political activity, for example, than disappointment-dissatisfaction. If these assumptions are correct, then responses to satisfaction questions are multidimensional and are embedded in a variety of psychological and behavioral processes. Without deeper probing into the meaning-context of superficially similar responses, satifaction is not measured at all. Or, to put it in an alternative and less drastic way, satisfaction is measured at such a low level of resolution that it is difficult or impossible to know exactly what is observed.

Levels of Development

There is another sense in which similar responses may obscure qualitative differences. Many theories of human developmental stages or levels exist; Maslow's need hierarchy is perhaps the most frequently cited. If one accepts

the validity of the developmental perspective and more particularly a theory of emotional development such as Drabowski and Piechowski's (1977), then the same degree of satisfaction registered by answers to a survey question may conceal vastly different levels of emotional maturity. Differences in level imply differences in behavior, and again, the power of resolution of the survey instrument is too low (comparable work on ego and moral development can be found in Loevinger, 1976, and Boyce and Jensen, 1978).

Change Versus Level

Within the complex set of measures of well-being used in SIWB is a subset of measures indicating perceptions of long term change. These measures exhibit a consistent lack of relationship to a heterogeneous set of more specific concerns. The authors conclude that: "It would appear that people are quite able to keep notions about *change* in well-being distinct from notions about its current *level*, and that their answers to questions about current levels are relatively uninfluenced by whatever may be their personal histories of past change, or expectations about future changes" (Andrews and Withey, 1976, p. 166).

This point is intriguing and relates directly to results of a survey I conducted some years ago (Wilcox, 1974) that started me pondering the nature of satisfaction questions. The survey focused on community satisfaction and included general questions on present satisfaction, changes in satisfaction over the past 5 years (past satisfaction), and expectations about changes over the next 5 years (future expectations). It also included a set of 16 more specific present-satisfaction questions (e.g., level of taxes, parks and playgrounds) and a comparable set of future expectation items (the present-satisfaction items overlap those analyzed in QOAL, chapter 7). Finally, it included a question on preferred community size and one on preferred growth rate, items that were the central policy-related focus of the survey.

Briefly, the pertinent results are the following: (1) parallel to SIWB, general present satisfactions related best to specific satisfactions, although general future expectations did almost as well, (2) both general past satisfaction and general future expectations related much more systematically to *specific* future expectations than did general present satisfaction, and (3) general past and general future were much more strongly related to preferred community size and preferred growth rate than was general present.

The last two results, in particular, suggest that perceptions about *change* in satisfaction (or happiness or well-being), depending on the analysis, may be more empirically relevant than perceptions about *level*. The assumption incorporated in QOAL and SIWB is that levels (and, more specifically, *present* levels) are the key aspect of quality-of-life concerns that need to be measured. This assumption appears to lack theoretical justification, and it is clearly debatable on empirical grounds.

LEVEL 2: PERCEIVED QUALITY OF LIFE

If we now assume, for the sake of exposition, that the questions of the adequacy and validity of measures of satisfaction have been satisfactorily answered, we can examine questions at our second level. Specifically, do these measures of satisfaction adequately reflect *perceived quality of life* (as typically phrased by SIWB)?

Emotional Experience

Asking people how satisfied they feel about various life concerns is taken here to be equivalent to asking them to report on their emotional experience. Does asking questions about satisfaction (or happiness or pleasure or delight) adequately capture the emotional experience of the respondents to the QOAL and SIWB surveys?[5] A few selected observations suggest that this may not be the case.

Few modern theories of emotion would suggest much utility in narrowing our emotional experience to a dimension of satisfaction, whether it be the biologically derived approach of Pugh (1977), the sociologically oriented theory of Kemper (1978), or the philosophically informed treatments of Solomon (1976) or Abelson (1977), among many others. Kemper's social-interactional theory of emotions separates them into three types—structural, antecedent, and consequent—depending on their location in ongoing power-status relationships, an approach difficult for studies like QOAL and SIWB to accommodate. Solomon, after developing a "logic of emotions," analyzes, in an admittedly exploratory effort, the following: anger, angst, anxiety, anguish, contempt, contentment (the closest to satisfaction), depression, despair, dread, duty, embarrassment, envy, faith, fear, friendship, frustration, gratitude, guilt, hate, hope, indifference, indignation, innocence, jealousy, joy, love, pity, pride, regret, remorse, resentment, respect, sadness, shame, vanity, worship. It would seem difficult to maintain that this or a similar range of emotions is encompassed by the D-T scale or the QOAL satisfaction scale.[6]

Personality

If respondents can be asked how they feel about a multitude of life domains and criteria, as they were in the surveys of QOAL and SIWB, they could also be asked about other perceptions of themselves, including perceptions of themselves as persons (normatively oriented conceptions of personality are surveyed

---

[5] The affective and cognitive components of responses to some of these questions is currently being investigated. See Andrews and McKennell (1980) and McKennell and Andrews (1980).

[6] Such theories might be further augmented by the concept of *levels of emotional development* mentioned above. Reviews of theories of emotions can be found in chapter 1 of Kemper (1978) and Izard (1977).

in Coan, 1977). Such questions might derive from the trait approach to personality theory, which in a recent selective compendium included achievement strivings, anxiety, authoritarianism, dogmatism, field dependence, introversion-extroversion, locus of control, machiavellianism, need for approval, power motive, repression-sensitivation, sensation seeking, and trust (London and Exner, 1978). It might derive from the holistic approach to personality, which would involve more complex interrogations in terms, say, of one or more of the conceptualizations compared in Salvatore Maddi's *Personality Theories* (1976). The possibilities could be amplified by roaming through the sources cited in Miller's *Handbook of Research Design and Social Measurement* (1977) as well as through the various volumes of *Advances in Psychological Assessment* and the *Annual Review of Psychology*.

If it is objected that survey respondents can or will not accurately assess, for example, how dogmatic they are, this objection may be well taken. However, we maintain that if it differs, it differs only in degree from accurate assessment of a respondent's satisfaction or happiness. If it is objected that such questions are not eliciting perceptions of quality of life, we must ask in what way personality questions differ from satisfaction questions. Are individuals' reports of how happy they feel any more an indication of perceived quality of life than reports of how dogmatic they are? A positive response to this query requires a theoretical/conceptual justification of the presence of a linkage between the concept perceived quality of life and satisfaction questions and the absence of any such linkage to personality questions. We haven't been able to uncover any such justification. Consequently, we must assume that QOAL and SIWB have not considered entire areas of self-perception that relate to their stated concerns.

Assessments of Quality

In our discussion of level 1, we suggested that different models of evaluation might lead to qualitatively different assessments of satisfaction that would go undetected by the standard satisfaction scales. According to some of these models, respondents would make judgments about their quality of life and then make a separate judgment about their degree of satisfaction with that assessed quality. In fact, pilot studies exist that "indicated that individuals did distinguish between perceptions of quality and judgments of satisfaction and that the quality judgments were more closely tied to objective indicators such as income than the satisfaction measures" (Atkinson, 1978, p. 3).[7]

In the paper from which this quotation is taken, the author observes "that the perceived quality of life is a multidimensional concept. Satisfaction is one of those dimensions but whether it is the most significant of them depends on the

[7]These studies are part of a Canadian research program comparable to that reported in QOAL and SIWB led by Tom Atkinson at the Institute for Behavioral Research, York University, Toronto, Canada. See also Atkinson (1977) on the point discussed in the text. The two programs are now doing cooperative research.

theoretical orientation of the reader" (Atkinson, 1978, p. 3). There is no disagreement on the point about multidimensionality. Assuming that satisfaction is one of those dimensions, the preceding two sections have essentially elaborated the nonexhaustive nature of the measures employed in QOAL, SIWB, and similar studies. Thus, the difficulty lies in singling out satisfaction measures without theoretical justification for their preeminence.

However, in this section we go one step further and question whether satisfaction is one of the dimensions of perceived quality of life at all. Asking people to assess the quality of their lives is, on its face, the way to obtain data on perceived quality of life. Asking them how satisfied they are with those rankings, ratings, or whatever is a conceptually distinct operation (as would be more difficult questions asking how personality traits might influence those judgments).[8] It is true that measures of quality of life and measures of satisfaction may be highly correlated under certain conditions, but this simply supports the potential indirect measurement validity of satisfaction measures, not their direct measurement (terminology from Payne, 1975).

We conclude that satisfaction measures are at best indirect indicators of perceived quality of life. If this argument is accepted, it appears to lead to one of three alternatives for future survey studies: (1) omit questions about satisfaction and simply ask about perceived quality of life directly, (2) omit the concept of perceived quality of life and ask satisfaction questions under some other rationale, or (3) ask both and investigate the relationships between the two.

LEVEL 3: QUALITY OF LIFE

If we make a second assumption, for the sake of exposition, that perceived quality of life has been adequately and validly measured by the questions of QOAL and SIWB, we can ascend to the third level of our critical odyssey. Is measuring how people perceive the quality of their lives the same as measuring the quality of their lives? Such a connection is implicitly and explicitly made at several places in QOAL and SIWB (although some ambiguity exists, to which we shall return).

After justifying a negative response to this question, we shall consider two additional questions that follow naturally in the context of QOAL and SIWB: (1) Are satisfaction and happiness the dimensions that define quality of life? (2) If not, what dimensions do define quality of life?

Perceptions and the Quality of Life

The stage for this section can be set by quoting from Nicholas Rescher's philosophically based discussion in *Welfare* (1972):

[8] Given the quote by Atkinson (1978), it is clear that this distinction has not gone unnoticed. We differ only in treating it as creating a much more serious problem for past and ongoing research.

> A person's welfare is a matter of objective conditions and circumstances, not of his subjective state. His material prosperity, his mental health, his physical condition—all these are objective facts that can be as well (or better) known to others as to himself: his family doctor, his financial counselor, his legal adviser, etc. For the state of his welfare—unlike any fleeting feelings of contentment or happiness or elation—is not a transient mental state of subjective psychological mood that the subject himself is, in the very nature of things, best fitted to judge. A person's welfare is determined by his state and condition in certain specifiable and overt respects: health, financial status, and the like. Judgments of welfare are matters of (objective) *knowledge* and not matters of (subjective) *feeling*. Thus whether a person's welfare is in better-or-worse-shape than it used to be is an issue about which others may well be better informed than he. Welfare is thus not in any immediate way a matter of psychological feelings or moods or states of mind; rather, it is a function of the extent to which certain objective circumstances are realized, namely, those generally regarded as representing requisites for the achievement of happiness in the existing life environment. Welfare hinges upon objectively determinable circumstances, and a sphere of action is consequently opened up for the effective workings of "outside" expertise. (p. 17)

Leaving until later the relationship of welfare to happiness, Rescher's position can be taken one step further. In addition to rendering much expertise irrelevant (including that of social scientists), the notion that welfare can be subjectively determined would also render most philosophical discourse on ethical and moral issues irrelevant. What would remain would not be much more than a combination of ethical relativism and metaethical emotivism. The same would hold true for values other than welfare that might be considered components of quality of life. The issues involved cannot be elaborated here, so we must simply state that we feel such a position to be unacceptable and untenable (Brandt, 1959, and Garner and Rosen, 1967, are useful introductions to the ethical and axiological issues involved).

Further, when the subjective results of surveys are aggregated into a measure of the quality of life of a nation, two additional and related assumptions are implied. The first is that "truth" can be established by a consensus based on processing of relatively immediate sense data. (Another way of putting this is to say that the surveys of QOAL and SIWB constitute examples of Lockean inquiring systems, to use Churchman's terminology, 1971.) Although such a simple epistemological approach may be valid in certain contexts for generating knowledge, the quality of a nation's life is not one of them. The second assumption is that the quality of life within a nation can be determined by a simple aggregation of the quality of life of individuals living within its boundaries. This in turn not only assumes equal weighting of individuals but also the absence of any structural (or systemic or global) dimension that might make the quality of life of the whole different from the quality of life of the parts.

Anyone working within a systemic perspective is likely to treat such an assumption as an example of individualistic fallacy and, to say the least, be skeptical of conclusions based on it (see Alker, 1965, p. 103).

Happiness and the Quality of Life

Many people consider happiness, as Rescher apparently does in the above quote, to be essentially equivalent to the presence of pleasure and the absence of pain. Although evidence is apparently not available, it is possible that some such hedonic calculus underlies the responses of most individuals to survey questions such as those emphasized in QOAL and SIWB.

The idea of happiness, however, has been treated by many philosophers (and nonphilosophers) since and including the Greeks as something much more involved than simple pleasure and pain. The various considerations need not be recounted here, for there are numerous accounts available in ethical theory. An excellent example is McGill's *The Idea of Happiness* (1967) where the similarities and differences among three broad nonpleasure approaches are ably presented (i.e., eudaemonism, self-realization theory, and self-actualization theory).

An alternative and frequently taken approach is to accept a pleasure-oriented definition of happiness but to dispute the desirability of happiness as a criterion of the good life. This is the tack taken by Kaufmann in *Without Guilt and Justice* (1973) where the need for alienation and the incompatibility of autonomy and happiness are central themes.[9] Both approaches are incompatible with assessing the quality of life of individuals by the extent to which they say they are happy and/or satisfied. Neither approach is compatible with taking as a standard to emulate the high levels of satisfaction one could register by surveying the inhabitants of *Brave New World* or *1984*.

Quality of Life Expanded

Happiness and satisfaction, of course, are not the only values or criteria by which one might assess quality of life. Bypassing the unnecessary attempt to compile an exhaustive list, we shall simply place our final point within the context of one category of values—the political.

It is obvious that one prime audience for information on quality of life in the United States (or in other countries) is public officials. One does not need to rehearse the history of the social indicators movement or tabulate government and government-sponsored research to unearth the obvious. Public officials and entire governments for that matter have an interest in the welfare of the citizenry, if for no other reason than as a means of social control. Information

---

[9] A poignant literary example of this theme can be found in George R. R. Martin's "A Song for Lya" (1976). The difficulties involved in operationalizing complex theoretical concepts are examined intensively by Scaff (1978).

on happiness and satisfaction, particularly if narrowly conceived, consequently might be of considerable interest.

Other values of a more directly political nature might and clearly often do concern holders of political and administrative office. Among such values one would expect to find the following: personal rights, collective preference, justice, equality, freedom, and the public interest. Each of these values (or principles) is the subject of an extensive theoretical, philosophical, and empirical literature (two of many useful studies of this material are Braybrooke, 1968, and Barry and Rae, 1975). Each implies a set of institutional and organizational arrangements. The quality of political life is in large part contingent upon the nature of these arrangements. All of this, as well as a more probing analysis of happiness, is essentially bypassed by QOAL and SIWB and similar research. We submit that complex judgments of quality of life cannot be made properly, whether by public officials or others, without incorporating assessments of the status of such political values.

SUMMARY

Retracing our sequence of levels, we can draw summary conclusions from the points we have made. First, to the extent that the arguments made in level 1 have merit, the work of QOAL and SIWB have not measured satisfaction with any degree of accuracy. Second, measures of satisfaction, however accurate, are only one of many possible types of subjective measures and, in any case, are not direct measures of perceived quality of life. Third, successful measurement of perceived quality of life, in turn, is not the same as successful measurement of quality of life, although the latter effort is apparently one of the primary goals of the research under review.

Should one conclude from this critique that the research reported in QOAL and SIWB and similar studies is valueless? We think not. The first generation of studies in any area of research is certain to contain flaws and often to raise questions more than provide answers. Beyond such truisms of the scientific enterprise, however, we can suggest ways to make succeeding generations of studies more productive and cumulative, and it is to this task that the remainder of the chapter is devoted.

WHAT IS TO BE DONE?

As we moved from level 1 to level 3 commentary, the questions raised became more far reaching. In like manner, attempts to respond adequately should become more difficult and complex as the levels are ascended.

Level 1

The semantic problems with the delighted-terrible scale can be resolved by a series of methodological studies that would, for example, empirically investigate

perceptions of appropriate antonyms and compare response distributions using different antonymic combinations. Similarly, the pros and cons of labeling and category-versus-magnitude scaling appear amenable to resolution through routine, if exacting, technical research.

Positivity bias (and perhaps negativity bias for some individuals and subpopulations) and lack of self-knowledge, on the other hand, are complex multidimensional phenomena that imply correspondingly complex efforts to investigate and compensate for in interpretations. As mentioned above, the strategy of multiple triangulation appears the one to follow. Survey responses would be interpreted in the light of data gathered through in-depth interviews, observations of nonverbal communications and other forms of behavior, analysis of written documents, role playing, and other types of experiments.[10] Clearly such a strategy, however executed, would be prohibitively expensive for large samples (national or otherwise). Yet, a set of carefully designed studies on smaller samples might provide critical information on the direction, extent, and varieties of distortion present in survey data. This information could then be used at least roughly to establish *parameters of distortion* that could be used to place survey results in perspective.

The comments made on models of evaluation and levels of development move us even deeper into theoretical as well as methodological concerns. If, as we suspect, the balance and nature of the cognitive, affective, and conative components (to use one familiar triad) of reported satisfaction differ depending on models and levels, then satisfaction cannot be measured adequately until these theoretical perspectives are woven into the research instrument. Again, less ambitious studies in terms of sampling might usefully probe these possibilities. The complicating factor, of course, is that the models and levels literatures are in various stages of development, and a once-and-for-all plugging of this material into satisfaction research is unlikely in the near future.

Finally, the observations on levels versus change are based upon one example of conceptual elaboration of satisfaction measurement. This is actually taken much further in SIWB, where the authors develop a fairly complex typology of measures of "global well-being." The point of our criticism was not that conceptual elaboration had been neglected but that one type of global measure (i.e., absolute, general, full range) became the centerpiece of the research without adequate theoretical or empirical justification. There are at least three types of response to this criticism: (1) provide such justification, (2) conduct research giving various types of measures equal treatment, a scientific parallel to equal protection of the law, or (3) more explicitly designate the research as concerned only with a particular type of measure.

---

[10]Examples of increased methodological self-conscious that would be useful in such an enterprise can be found in Alker (1975), Brenner et al. (1978), Lofland (1976), and Schwartz and Jacobs (1979).

Levels 2 and 3

What could or should be done at levels 2 and 3 involves a potentially vast program of research, one that has already been hinted at and cannot be taken much further here. However, narrowing our focus temporarily, the relevance of criticisms and suggestions offered at these levels to the concerns of research such as that reported in QOAL and SIWB is clearly tied to the claims made for and by that research. To illustrate, we shall briefly traverse a series of possible claims that might be made of this research, backtracking momentarily to level 1. One possible claim is that the research applies specifically and exclusively to the use of the D-T scale (or one of the other indices or scales), a narrow operationism made notorious by the statement that intelligence is what intelligence tests test.

A second possible claim is that it applies to the measurement of general satisfaction. In this case, conceptual and methodological development is necessary so that one has some basis for judging when satisfaction has been adequately measured. Many of the considerations discussed in level 1 are applicable to such a claim.

A third type of claim, implied by the title of SIWB, is that the research applies to well-being or perceived well-being. This implies a much broader scope than satisfaction and relates to the first two points made in level 2 and to the second two in level 3.

A fourth claim, implied by the subtitle of SIWB and by the complete title of QOAL, is that perceived life quality is the focus of the research. Although this appears to capture much of the intent of both studies, our discussion of assessments of quality in level 2 suggests that the intent has not been realized.

The fifth claim, of course, is that the quality of life is the focus of the research. Although this claim is less explicit and prevalent, passages in both QOAL and SIWB imply it. Our discussion in level 3 suggests that this research is not even close to supporting such a claim, a point we take up again below.

In sum, what we have in QOAL and SIWB is considerable ambiguity. It is often not clear what claims are being made, for the terminology changes from one section to another, and there is not enough conceptual clarification to be sure what is intended. Stronger emphasis on conceptual explication is needed in future work of this kind. In our own estimation, the research discussed here comes closest to supporting a claim of useful exploratory work into perceptions of satisfaction, happiness, and other types or aspects of well-being. To what extent it can or will be linked to or expanded into research supporting broader claims is a question for the future.

Quality of Life

Whatever quality of life may constitute in actuality, one feels from reading these reports that the authors believe their research is relevant to its assessment.

The concluding sections of QOAL are particularly pertinent here. After presenting an objection similar to the quotation by Rescher discussed above, the authors comment:

> Thus direct information on actor satisfactions is merely one ingredient among many to be folded into the development of enlightened policy. Over the period during which we planned and carried out this study, we received numerous and sometimes vehement communications from colleagues who were concerned that in focusing a major study on the feelings people had about the major terms of their lives we would somehow mistake the perceived quality of their lives for its actual quality, as might be determined by some outside and presumably higher wisdom. While the details varied across these well-intentioned cautions, fears that we might foster the most superficial and individualistic "hedonism" as a prime guide for policy were commonly expressed.
>
> If an interest in how actors evaluate the situations in which they find themselves can be caricatured as "hedonism," then many of the concerns expressed in the Seashore essay can equally well be caricatured, although quite unfairly to our mind, as "paternalism." The actors themselves cannot be trusted to see their immediate situations in the sufficiently broad perspective available to intellectuals and decisionmaking elites and hence, "papa knows best" what policy thrusts would really serve to promote the general weal. (p. 503)

The authors go on to suggest that neither position is worth extended scrutiny and that a middle course must be chosen. We would construe the problem somewhat differently and no doubt controversially. Judging quality of life, whether generally or specifically, is a task that inherently resides in a complex process of philosophical, theoretical, and empirical discourse whose primary participants are indeed experts. This process is in a primitive stage of development. Little consensus has emerged over what characterizes a high quality of life, particularly in its more general and abstract manifestations. (This lack of consensus is emphasized in EPA, 1973, and has led to a "minimalist" approach in Markley and Bagley, 1975, as a way of partially circumventing disagreement.) Indeed, little consensus has emerged over the identity of the experts, although one might speculate that a complex configuration of philosophers, scientists, and humanists would be a place to start. (An example of a discussion of part of such a process can be found in MacRae, 1976. Advocacy of more politically oriented processes is contained in de Neufville, 1978-1979, and Ackoff, 1975.) Granting all of this amorphousness, interrogation of the general public cannot be substituted for such a process and is necessarily secondary. We would submit that this position is not paternalistic at all, no more than the position that the debate over competing theories about the origin of the universe is best left to cosmologists and other experts and not consigned to a plurality determination by a vote of survey respondents.

However, it is indeed the case that members of the general public can become expert or close to it in at least some areas of quality-of-life assessment. For example, judging the quality of one's physical/mental health potentially could be (and in some cases is) done within limits by nonprofessionals trained in symptomology and diagnostic procedures. Indeed, we would not only advocate perfecting a process of quality-of-life assessment (that needs indefinite perfecting) but increasing the spread and level of such expertise throughout society to the fullest extent possible (including the expertise involved in choosing among experts).

CONCLUSION

In this chapter we have expressed a series of dissatisfactions with quality-of-life research based on measures of satisfactions or similar items, particularly as represented by *The Quality of American Life* and *Social Indicators of Well-Being*. Many of the problems we have discussed have been recognized, at least in passing, by the authors of these pathbreaking works. In the interest of breaking the path even further, we have tried to etch our discontents in as bold relief as possible.

Thorough criticism of research in this area is particularly important because of its potential policy applications. The net effect of some of the distortions (particularly positivity bias and lack of self-knowledge) is to register levels of satisfaction far above what they actually are. Aware of parts of this problem, the authors of QOAL propose emphasizing relative levels of satisfaction (among groups, among domains, etc.). However, policymakers, if not academics, are likely to overlook such niceties and focus to some extent on reported levels of satisfaction in an absolute sense.

Similarly, policymakers might tend to focus on "general measures of satisfaction" or "global measures of well-being" when more specific measures may be those with explanatory power. The importance of the generality-specificity distinction has been repeatedly stressed. Higher correlations are obtained when more specific items are related to specific behavior in attitude studies (Mervielde, 1977, p. 262; Rosenberg, 1979, p. 279). Different patterns of relationships are found when general and specific housing satisfaction items are used in the QOAL study (pp. 124-128). In my own research, several specific community satisfaction items were as strongly or more strongly related even to *general* questions about community size than was the general satisfaction item.

Going a couple of steps further, Barbash's (1976) review of job satisfaction attitude surveys contains the following assessment:

> Is job satisfaction a unitary state of mind so that the common survey question—"All in all, how satisfied would you say you are with your job"—is likely to represent a calculation, however imprecise, which totes up costs and benefits? There is no consensual answer on this one. There *is*

> a consensus that a general question of satisfaction or discontent has little operational meaning, which can only be derived from job satisfaction perceived as multi-dimensional or multi-faceted. (p. 17)

Finally, a study of quality of life in new communities included a question asking respondents what the phrase "quality of life" meant to them. Seventeen percent included contentment, well-being, happiness in their responses, but five other more specific components, headed by economic security, registered higher percentages (Zehner, 1977, pp. 24-25).

The differential relevance of general and specific questions highlights the care with which data on subjective social indicators must be analyzed and interpreted. The significance of global measures, which have received the most emphasis, is highly problematic and policies based on such information are even more problematic.

In spite of such cautions, research on subjective social indicators and, more narrowly, on subjective indicators of quality of life is an encouraging sign of the concern with a wider range of human values in the social sciences and in political life. Highlighting the problems in this research will hopefully encourage a comparably wide-ranging process that will partially transform our fragmented intellectual enterprise into the interwoven dialogue needed for so central a concern as the quality of life, both human and otherwise.

REFERENCES

Abelson, Raziel. 1977. *Persons: A Study in Philosophical Psychology.* New York: St. Martin's Press.

Ackoff, Russell L. 1975. "Does Quality of Life Have to Be Quantified?" *General Systems.* vol. 20, 213-219.

Alker, Hayward R., Jr. 1965. *Mathematics and Politics.* New York: Macmillan.

____. 1975. "Polimetrics: Its Descriptive Foundations." In Fred J. Greenstein and Nelson W. Polsby, eds., *Handbook of Political Science,* vol. 7, 139-210. Reading, Mass.: Addison-Wesley.

Allardt, Erik. 1978. "The Relationship Between Objective and Subjective Indicators in the Light of a Comparative Study." In Richard F. Tomasson, ed., *Comparative Studies in Sociology,* 203-216. Greenwich, Conn.: J.A.I. Press.

Andrews, Frank M., and Ronald F. Inglehart. 1979. "The Structure of Subjective Well-Being in Nine Western Societies." *Social Indicators Research* 6 (January): 73-90.

Andrews, Frank M., and Aubrey C. McKennell. 1980. "Measures of Self-Reported Well Being: Their Affective, Cognitive, and Other Components." *Social Indicators Research* 8: 127-155.

Andrews, Frank M., and Stephen B. Withey. 1976. *Social Indicators of Well-Being: Americans' Perceptions of Life Quality.* New York: Plenum Press.

Atkinson, Tom. 1977. "Is Satisfaction a Good Measure of the Perceived Quality

of Life?" Paper presented at the 1977 meetings of the American Statistical Association, August 1977.

____. 1978. "Trends in Life Satisfaction Among Canadians, 1968-1977." Paper presented at the IXth World Congress of Sociology, Uppsala, Sweden, August 1978.

Barbash, Jack. 1976. *Job Satisfaction Attitudes Surveys.* Paris: Organization for Economic Cooperation and Development.

Barry, Brian, and Douglas W. Rae. 1975. "Political Evaluation." In Fred J. Greenstein and Nelson W. Polsby, eds., *Handbook of Political Science,* vol. 1, 337-401. Reading, Mass.: Addison-Wesley.

Boyce, William D., and Larry Cyril Jensen. 1978. *Moral Reasoning: A Psychological-Philosophical Integration.* Lincoln: University of Nebraska Press.

Bradburn, Norman M., Seymour Sudman, and Associates. 1979. *Improving Interview Method and Questionnaire Design.* San Francisco: Jossey-Bass.

Brandt, Richard. 1959. *Ethical Theory.* Englewood Cliffs, N.J.: Prentice-Hall.

Braybrooke, David. 1968. *Three Tests for Democracy: Personal Rights, Human Welfare, Collective Preference.* New York: Random House.

Brenner, Michael, et al., eds. 1978. *The Social Contexts of Method.* New York: St. Martin's Press.

Campbell, Angus, Philip E. Converse, and Willard L. Rodgers. 1976. *The Quality of American Life: Perceptions, Evaluations, Satisfactions.* New York: Russell Sage Foundation.

Churchman, C. West. 1971. *The Design of Inquiring Systems.* New York: Basic Books.

Clark, Herbert H., and Eve V. Clark. 1977. *Psychology and Language: An Introduction to Psycholinguistics.* New York: Harcourt Brace Jovanovich.

Coan, Richard W. 1977. *Hero, Artist, Sage, or Saint?* New York: Columbia University Press.

Crosby, Faye. 1979. "Relative Deprivation Revisited: A Response to Miller, Bolce, and Halligan." *American Political Science Review* 73 (March): 103-112.

Dabrowski, Kazimierz, and Michael M. Piechowski. 1977. *Theory of Levels of Emotional Development,* vol. 1. Oceanside, N.Y.: Dabor Science Publications.

de Neufville, Judith Innes. 1978-1979. "Validating Policy Indicators," *Policy Sciences* 10: 171-188.

Denzin, Norman K. 1970. *The Research Act.* Chicago: Aldine.

Garner, Richard T., and Bernard Rosen. 1967. *Moral Philosophy.* New York: Macmillan.

Grier, William H., and Price M. Cobbs. 1968. *Black Rage.* New York: Basic Books.

Gur, Ruben C., and Harold A. Sackheim. 1979. "Self-Deception: A Concept in Search of a Phenomenon." *Journal of Personality and Social Psychology* 37 (February): 147-169.

Izard, Carroll E. 1977. *Human Emotions.* New York: Plenum Press.

Kaufmann, Walter. 1973. *Without Guilt and Justice: From Decidophobia to Autonomy.* New York: Peter H. Wyden.

Kemper, Theodore D. 1978. *A Social Interactional Theory of Emotions.* New York: Wiley.

Leege, David C., and Wayne L. Francis. 1974. *Political Research.* New York: Basic Books.

Lodge, Milton, and Bernard Tursky. 1979. "Comparisons Between Category and Magnitude Scaling of Political Opinion Employing SRC/CPS Items." *American Political Science Review* 73 (March): 50-66.

Loevinger, Jane. 1976. *Ego Development: Conceptions and Theories.* San Francisco: Jossey-Bass.

Lofland, John. 1976. *Doing Social Life.* New York: Wiley.

London, Harvey, and John E. Exner, Jr., eds. 1978. *Dimensions of Personality.* New York: Wiley.

Lyons, John. 1977. *Semantics,* vols. 1 and 2. Cambridge, England: Cambridge University Press.

MacRae, Duncan, Jr. 1976. *The Social Function of Social Science.* New Haven: Yale University Press.

McGill, V. J. 1967. *The Idea of Happiness.* New York: Frederick A. Praeger.

McKennell, Aubrey C., and Frank M. Andrews. 1980. "Models of Cognition and Affect in Perceptions of Well-Being." *Social Indicators Research* 8: 257-298.

McReynolds, Paul, ed. 1977. *Advances in Psychological Assessment.* San Francisco: Jossey-Bass.

Maddi, Salvatore R. 1976. *Personality Theories: A Comparative Analysis,* 3d ed. Homewood, Ill.: Dorsey Press.

Markley, O. W., and Marilyn Bagley. 1975. *Minimum Standards for Quality of Life.* Menlo Park, Calif.: Stanford Research Institute.

Martin, George R. R. 1976. "A Song for Lya," *A Song for Lya and Other Stories.* New York: Avon Books.

Mervielde, Ivan. 1977. "Methodological Problems of Research About Attitude-Behavior Consistency." *Quality and Quantity* 11: 259-281.

Miller, Delbert C. 1977. *Handbook of Research Design and Social Measurement,* 3d ed. New York: David McKay.

Miller, George A., and Philip N. Johnson-Laird. 1976. *Language and Perception.* Cambridge, Mass.: Harvard University Press.

Payne, James L. 1975. *Principles of Social Science Measurement.* College Station, Tex.: Lytton.

Phillips, Derek L. 1971. *Knowledge from What?* Chicago: Rand McNally.

Pugh, George Edgin. 1977. *The Biological Origin of Human Values.* New York: Basic Books.

Rescher, Nicholas. 1972. *Welfare: The Social Issues in Philosophical Perspective.* Pittsburgh: University of Pittsburgh Press.

Rosenberg, Morris. 1979. *Conceiving the Self.* New York: Basic Books.

Scaff, Lawrence A. 1978. "Conceptualizing Alienation: Reductionism and the Problem of Meaning," *Philosophy of the Social Sciences* 8: 241-260.

Schneider, Mark. 1976. "The 'Quality of Life' and Social Indicators Research." *Public Administration Review* 36: 297-305.

Smith, H. W. 1975. *Strategies of Social Research.* Englewood Cliffs, N.J.: Prentice-Hall.

Smith, Tom W. 1979. "Happiness: Time Trends, Seasonal Variations, Inter-survey Differences and Other Mysteries." *Social Psychological Quarterly* 42(1): 18-30.

Solomon, Robert C. 1976. *The Passions.* Garden City, N.Y.: Doubleday Anchor Press.

Stipak, Brian. 1979. "Citizen Satisfaction with Urban Services: Potential Misuse as a Performance Indicator." *Public Administration Review* 39: 46-52.

Sudman, Seymour, and Norman M. Bradburn. 1974. *Response Effects in Surveys: A Review and Synthesis.* Chicago: Aldine.

U.S. Environmental Protection Agency. 1973. *The Quality of Life Concept: A Potential New Tool for Decision-Makers.* Washington, D.C: U.S. Government Printing Office.

Wilcox, Allen R., et al. 1974. *Blue Ribbon Task Force Program Report No. 9: Optimum Size and Psychology of Growth.* Washoe County, Nev.: Area Council of Governments.

Williams, Robin M., Jr. 1975. "Relative Deprivation." In Lewis A. Coser, ed., *The Idea of Social Structure: Papers in Honor of Robert K. Merton.* New York: Harcourt Brace Jovanovich.

Zehner, Robert B. 1977. *Indicators of the Quality of Life in New Communities.* Cambridge, Mass.: Ballinger.

Comments

Frank M. Andrews
Institute for Social Research
University of Michigan

INTRODUCTION

Wilcox's chapter expresses a range of concerns about the goals and procedures adopted in two studies of perceived life quality. These studies were conducted at the University of Michigan's Institute for Social Research and are reported in two widely cited books: *Social Indicators of Well-Being: Americans' Perceptions of Life Quality*, abbreviated by Wilcox as SIWB (Andrews and Withey, 1976) and *The Quality of American Life: Perceptions, Evaluations, and Satisfactions* (QOAL) (Campbell, Converse, and Rodgers, 1976).

In this commentary on Wilcox's chapter, I discuss three things: (1) in order that Wilcox's comments may be understood in an appropriate context, I discuss some of the developmental and practical considerations that influenced these studies; (2) I elaborate on one of the key conceptual issues raised by Wilcox's paper—the need for and the usefulness of multidimensional assessments of life quality, and I report some recent empirical results we have developed that bear on this topic; and (3) I note some of the more significant areas of agreement and disagreement between Wilcox and myself.

The casual reader of Wilcox's paper will perhaps be struck by the fractious tone of the presentation and the range of critical comments he makes about some of the work reported in the SIWB and QOAL books. Yet the casual reader may miss what I believe are the more positive purposes that motivate Wilcox's paper and the forward-looking suggestions that he offers. It should be emphasized that Wilcox himself apparently supports the basic undertakings of the studies reported in SIWB and QOAL, for he writes that "research on subjective social indicators and, more narrowly, on subjective indicators of quality of life, is an encouraging sign of the concern with a wider range of human values in the social sciences and in political life." Furthermore, Wilcox describes the SIWB and QOAL books as "clearly pathbreaking and sophisticated in design, analysis,

I am grateful to Angus Campbell, Willard Rodgers, and Stephen Withey for their helpful comments on a draft of these comments.

and interpretation," and as presenting "useful exploratory work into perceptions of satisfaction, happiness, and other types or aspects of well-being." That this pair of pioneering studies in a large, new, and important area are accorded such comments suggests that the studies have achieved their basic purposes. (Beyond the "exploratory work" mentioned by Wilcox, however, I believe the books are also important for the broad and basic descriptive data about evaluations of well-being by representative national samples of American adults in the first part of the 1970's. This component of the work Wilcox does not discuss. However, this descriptive material has already been used by other investigators to compare their own local data with national figures, and with the passage of time, these descriptive national data seem likely to increase in historical importance as baselines for assessing future changes.)

Thus, lest the critical presentation style of Wilcox's paper mislead the casual reader, I would like to suggest that Wilcox's comments can most usefully be viewed not as a pugnacious attempt to flail at a pair of studies that have accomplished much but rather as a series of research suggestions for broadening and deepening our understandings about perceptions of life quality. The reasonableness of such a perspective becomes clear when one takes into account some of the developmental and practical considerations that influenced SIWB and QOAL.

DEVELOPMENTAL AND PRACTICAL CONSIDERATIONS

Although I believe that many of Wilcox's observations about work reported in SIWB and QOAL are correct, the implications of his observations depend on an understanding of the developmental role these studies have played in the social indicators movement and on some of the practical considerations of actually conducting empirical survey research on moderate-size representative national samples of Americans. It is useful to briefly describe the context in which these studies were carried out and the milieu in which their results now exist.

It is important to realize that these books report the first major studies of their kind in what was (in the early 1970's) a new field of inquiry. As the work was being done, many conceptual and methodological issues arose for which there were few if any precedents to provide ready answers. However, it seemed desirable to push ahead in the best way available at the time and to try to develop a significant body of basic research knowledge. This pushing ahead was not a slapdash process. Both studies include unusually extensive attention to methodological issues, and many of the methods and procedures used in the studies were adopted only after considerable investigation. (These investigations are reported at length in SIWB and QOAL, and many readers seem to regard these portions of the books as particularly interesting and valuable.)

As a general principle, I suggest that it is often more productive in the long run to move ahead and produce research results that subsequently can be assessed, revised, and extended than to become bogged down in what might become

endless debates. This is not to deny the importance or interest of theoretical debates or that they can and should continue, but they are likely to be more fruitfully pursued in the presence of relevant empirical results and in the light of the experiences gained in developing and applying those results.

Not only is it important to avoid becoming mired in unresolvable issues, one also must realize that the scope of a particular research undertaking must be confined to what is feasible and manageable with available resources. Experienced researchers are well aware that it is frequently impossible to explore in a single series of studies all of the potentially interesting and relevant avenues of inquiry that may radiate from a particular topic.

When these two considerations are acknowledged, many of the topics discussed by Wilcox become agenda for future research rather than cogent criticisms of the studies reported in SIWB and QOAL. Let us consider several specific examples.

1. In the subsection, "Emotional Experience," Wilcox observes that there are a wide range of emotional experiences that might have some relevance for people's sense of well-being. He goes on to cite a list of 36 such emotions, ranging from anger and angst to vanity and worship. Then he observes that "it would seem difficult to maintain that this or a similar range of emotions is encompassed" by the scales used in SIWB and QOAL. Of course, he is correct, but the point is specious: Neither SIWB nor QOAL set out to provide a full-blown description of emotional experiences, and neither book claims to do so. Furthermore, it would have been impossible to accomplish such a task in any reasonable degree of depth within realistic interviewing time constraints. Is Wilcox actually suggesting that these studies could or should have given detailed attention to the 36 emotions he lists? His text is ambiguous on this point, but let us presume that he recognizes that such an undertaking would have been infeasible for these studies in the presence of their other commitments. Then this subsection of his paper becomes a suggestion that future studies of perceived well-being devote some resources to exploring linkages between various emotions and well-being. With this I would agree.

2. As a second example of the need to interpret Wilcox's observations in the light of developmental and practical considerations, we have only to turn to the immediately following subsection, "Personality." Here Wilcox observes that respondents could have been asked about their perceptions of themselves as persons; then he mentions more than a dozen traits, including field dependence, machiavellianism, and power motive; and he concludes that "QOAL and SIWB have not considered entire areas of self-perception that relate to their stated concerns." Again, he is correct, and again his observation seems more appropriate as a suggestion for future research than as a criticism of SIWB or QOAL. Are we to believe that Wilcox thinks SIWB or QOAL should have provided an in-depth study of personality as it relates to perceptions of life quality? Its infeasibility in a pioneering study of well-being is obvious, and to suggest that these studies should have explored both a significant number of personality

traits and a wide range of emotions would be utterly naive. Clearly, we must interpret Wilcox's observations as suggesting research agendas for the future.

3. As a third example of the need to interpret Wilcox's comments in an appropriate contextual framework, consider his discussion of what he calls the claims made by SIWB and QOAL in regard to studying the quality of life (in his subsection titled "Levels 2 and 3"). He chides the books for being "ambiguous," and for not presenting enough "conceptual clarification to be sure what is intended." In actuality, anyone who takes the time to examine the books can find out what was done and what was learned—and that is what was intended. The term *quality of life* is obviously a broad and imprecise one, and debate about its meaning and limits has been going on for millennia and will continue. This is an issue in which one could become mired, and I suggest that one of the strengths of SIWB and QOAL is that they did not try to discover or say all that might be relevant to quality of life. Clearly, however, both SIWB and QOAL have some relevance to the topic of life quality, and the inclusion of this phrase in their titles seems reasonable. I think Wilcox's comments on this topic are best interpreted as reminders that further debate and refinement of terms such as quality of life and well-being are possible.

The three examples above illustrate why I believe that many of the points made by Wilcox, which are ostensibly presented as criticism of SIWB or QOAL, are not really relevant as serious suggestions for ways those studies should have been done differently but rather are most useful as suggestions for topics that might be explored in the future. And because Wilcox's comments are by no means the only responses to the studies reported in SIWB and QOAL, it seems appropriate to conclude this section of my commentary by briefly noting certain other features of the past and current scientific context out of which the studies emerged and into which their results have been disseminated.

As noted previously, both studies originated in the Institute for Social Research at the University of Michigan. They were conducted by two separate research teams operating independently, although they remained informed of each other's work and shared certain premises about goals and procedures. The teams were funded from separate sources[1] and had somewhat different interests and ideas about ways to proceed. Although the two books that resulted are distinctly different—each one covering territory not covered by the other—the basic results, where they can be directly compared, are highly consistent. This replicability of the results is a very important general finding from this body of work. The fact that each study supports the other greatly enhances the confidence that can be placed in each one individually.

Since their publication in 1976, both SIWB and QOAL have been prominently (and in general favorably) reviewed in the major relevant journals (Burke,

[1] SIWB reports investigations that were funded mainly by the National Science Foundation. The main support for the work reported in QOAL was provided by the Russell Sage Foundation.

1979a, 1979b; Cohn, 1977; Glock, 1976; Land, 1978; Mason, 1978; Phillips, 1978). Perhaps more important, the methods and findings described in these books and various journal articles derived from the same studies (e.g., Andrews, 1974; Andrews and Crandall, 1976; Andrews and Inglehart, 1978; Andrews and Withey, 1974; Campbell, 1976; Rodgers, 1977; Rodgers and Converse, 1976) seem to have been of significant interest and use to other investigators studying perceptions of well-being. Other studies that have used directly or adapted the methods, questionnaire items, and/or response scales from our studies include those by Atkinson (1977), Butler (1978), Forti (1979), Headey (1980), McKennell et al. (1980), Michaelsen et al. (1976, 1980), Sontag et al. (1979), Vaughan and Lancaster (1979), Wilkening and McGranahan (1978), and Zautra et al. (1976).

ON THE MULTIDIMENSIONALITY OF ASSESSMENTS OF WELL-BEING

One of the most interesting topics raised by Wilcox concerns the multidimensional character of assessments of well-being. This is one of his central concerns—that there may be dimensions relevant to life quality other than or in addition to the satisfaction dimension. There is absolutely no debate about this. Both SIWB and QOAL include examples of other dimensions, including happiness, positive affect, negative affect, perceptions of change, desires for change, and current mood, and neither book implies that the life quality concept or even the narrower notion of perceived well-being is unidimensional.

Recent work by McKennell and me (Andrews and McKennell, 1980; McKennell, 1978; McKennell and Andrews, 1980) provides statistical evidence that the well-being measures explored in SIWB and in British studies conducted by Abrams (1973) and Hall (1976) seem to reflect at least two types of underlying components—cognition and affect. As we expected, satisfaction measures (and here the term refers to questions that ask specifically about satisfaction rather than in the generic sense used by Wilcox) seemed to reflect relatively more cognition than did happiness measures, which were relatively saturated with affect. However, in these analyses it did not appear that satisfaction measures were pure measures of cognition or that happiness measures reflected only affect.

Although there is no debate about the fact that multiple underlying dimensions can be identified in perceptions of well-being, there are important unresolved issues regarding how many assessment dimensions it is useful to distinguish. It seems likely that there is no single right answer here but that what proves useful depends on one's purposes and interests.

An analogy may help to clarify the issue. Consider the size of an automobile. For some purposes it is important to distinguish carefully between such size-related factors as passenger capacity, weight, engine horsepower, and fuel consumption. However, all these dimensions tend to vary together (the covaria-

tion is substantial but not perfect), and sometimes it is useful and appropriate to use a consolidated concept and think of cars as arrayed along a single size dimension ranging from small to large. This dimension is, in fact, in common use and demarcated by the advertisers with terms such as subcompact, compact, midsize, and fullsize. Of course, using the consolidated size concept for one purpose does not preclude using the more detailed dimensions, such as passenger capacity, for other purposes.

Now with respect to quality of life, I suggest that assessments of perceived well-being can be considered at varying degrees of consolidation, just like size-related characteristics of cars. At one extreme is the single conglomerate that summarizes many potentially distinguishable underlying dimensions.[2] It can be argued that such a consolidation is more than a mere statistical artifact: It taps something that people do experience. When a person takes action intended to change his or her level of general well-being, for example, migrating to a new area, taking a new job, or moving out of a parent's home or into a retirement center, it is presumably because on some consolidated assessment dimension well-being is expected to be improved.

Some of the measurement scales that are described in SIWB—the delighted-terrible, faces, circles, and ladder scales—yield measures that indicate well-being along a highly consolidated dimension. It turns out that the satisfaction scale used extensively in QOAL and to a lesser extent in SIWB also seems to reflect rather consolidated judgments about well-being.[3] The assessment of well-being along these rather consolidated dimensions proved useful in these early studies, but it in no way precludes the conceptualization of well-being as a multidimensional concept or the measurement of separate underlying dimensions.

The times when it will be useful to measure more detailed dimensions of well-being are when one suspects or can show that their relationships to other concepts of interest differ in significant and important ways and/or if one can identify specific dimensions of well-being that bear little statistical relationship to one another. Developing this body of knowledge has begun (e.g., SIWB, QOAL, McKennell, 1978) but is not yet very far advanced. Some of Wilcox's observations about SIWB and QOAL are in effect suggestions that future

---

[2] This idea of a consolidated assessment dimension should not be confused with the concept of global well-being as this term is used by Wilcox and SIWB and QOAL. Global well-being refers to *something that is being assessed* (life as a whole) and is an alternative to assessments of more specific life concerns such as housing, job, and so on. In contrast, the concept of consolidated assessment dimension refers to a characteristic of the *dimension* along which the assessment is made.

[3] In analyses reported in Andrews and McKennell (1980), we show that measures based on delighted-terrible scales reflected affect and cognition in about equal proportions and that measures derived from satisfaction scales, while somewhat more cognitively oriented, also reflected significant amounts of affect. Although these findings were not available when SIWB was written, we had done a series of analyses on measures of well-being and found that measures derived from answers to the delighted-terrible scale tended to occupy a central core position in the multidimensional space defined by the 60-plus measures we are examining (SIWB, chapter 2). This central position, of course, is in accord with the observation here that delighted-terrible scale measures (among others) reflect a highly consolidated assessment of well-being.

research examine how certain detailed dimensions of well-being relate to one another and to other variables of interest, and on this matter he and I completely agree. However, in addition, I suspect that there is likely to be a continuing need and usefulness for assessments of well-being along highly consolidated dimensions. I do not see that Wilcox has specifically addressed this matter.

SOME POINTS OF AGREEMENT AND DISAGREEMENT

The preceding sections of this commentary have noted various important areas where I am in general agreement with the ideas presented in Wilcox's chapter, and these will not be repeated here. So far as I am aware, he has not made factual errors in his statement about SIWB or QOAL. To the extent that we differ in our views, I believe these differences relate to the implications of his comments or to matters of judgment as to where the most fruitful and important lines of future investigation lie.

For example, his comments about the delighted-terrible scale (in the subsection "The Delighted and the Terrible") seem factually correct, but I believe he neglects to consider the fact that this scale was intended to yield highly consolidated assessments, as described above, and that as such a strict application of semantic logic may not be a relevant criterion. The scale has worked well in the face-to-face interview situation for which it was designed (the delighted-terrible scale format does not, however, seem well suited to telephone interviews), and it seems unlikely that major improvements can be achieved, though modest refinements may be possible and would be welcomed. If Wilcox or others wish to try making improvements, they are certainly invited to do so. Evidence of achieved improvement, however, is more useful than a mere discussion of possible weaknesses. Similar comments apply to many of Wilcox's other observations: What would be most helpful would be actually to demonstrate that something better can be done.

There are two areas, however, in which I have serious misgivings about what Wilcox has said, and these deserve special attention. The first of these involves his comments about positivity bias in the conclusion to his chapter. He writes: "The net effect of some of the distortions . . . is to register levels of satisfactions far above what they actually are." Everything that I have seen, including the detailed examination of this issue in QOAL, leads me to believe that that is unlikely to be true. On the contrary, given all the ways people change their lives and/or their criteria of judgment in order to enhance their sense of well-being, it would be surprising if large proportions of Americans were fundamentally dissatisfied with most of the common and basic life concerns that are the topics of SIWB and QOAL.

This assertion does not imply that life is heavenly for most people—surely it is not. However, there is an important psychological need to experience at least some positive elements in aspects of life that are closely linked to oneself, and

it appears that most people achieve this for themselves most of the time. One way of achieving it is to make changes in the objective circumstances of one's life, take a part-time job, move to a warmer climate, and so on. Another way is to adjust the criteria by which well-being is judged, for example, to decide that the way one presently spends one's days is good or to conclude that snow is fun or pretty. Both of these ways of coping with potential dissatisfactions are common, and the achievement of an improved sense of well-being by either means is real and important. If people then report substantial levels of satisfaction—as most people do—this seems likely to be a reflection of what they actually experience. At the very least, I believe that Wilcox should present some significant empirical evidence before he makes important assertions as sweeping as those about the extent of positivity bias present in these data.

The second and related area in which I feel Wilcox may be misleading his readers is his discussion of the relevance of perceptions for understanding life quality (in his subsections "Perceptions and the Quality of Life" and "Quality of Life"). Although I understand that Wilcox himself believes people's assessments of their own well-being are relevant to their sense of life quality (in accord with portions of his text quoted in the introductory section of this commentary), he devotes more space to a lengthy quotation from Rescher (to the effect that judgments of welfare are not matters of subjective feeling) and to an apparent endorsement of that view. While one might resolve this possible contradiction by applying a narrow definition of welfare, that is, by suggesting that Rescher's and Wilcox's use of the term welfare is distinct from what is implied by quality of life, that leaves the relevance of perceptions of well-being unaddressed. Neither SIWB nor QOAL state that subjective assessments of well-being are the only relevant aspects of life quality. One of their premises, however, is that people's feelings about their own well-being do matter and matter very much and that any attempt to assess life quality that neglects people's assessments of their own well-being must be incomplete.

REFERENCES

Abrams, Mark. 1973. "Subjective Social Indicators." *Social Trends*, no. 4, London: Her Majesty's Stationery Office.

Andrews, Frank M. 1974. "Social Indicators of Perceived Life Quality." *Social Indicators Research* 1: 279-299.

Andrews, Frank M., and Ronald F. Inglehart. 1978. "The Structure of Subjective Well-Being in Nine Western Societies." *Social Indicators Research* 6: 73-90.

Andrews, Frank M., and Aubrey C. McKennell. 1980. "Measures of Self-Reported Well-Being: Their Affective, Cognitive, and Other Components." *Social Indicators Research* 8: 127-155.

Andrews, Frank M., and Stephen B. Withey. 1974. "Developing Measures of Perceived Life Quality: Results from Several National Surveys." *Social Indicators Research* 1: 1-26.

____. 1976. *Social Indicators of Well-Being: Americans' Perceptions of Life Quality*. New York: Plenum Press.

Atkinson, Tom. 1977. "Is Satisfaction a Good Measure of the Perceived Quality of Life?" *Proceedings of the Social Statistics Section of the American Statistical Association*. Washington, D.C.: American Statistical Association.

Burke, Ronald J. 1979a. Review of *The Quality of American Life*. *Social Indicators Research* 6: 487-490.

____. 1979b. "Doing Better but Feeling Worse." *Contemporary Psychology* 24: 180-181.

Campbell, Angus. 1976. "Subjective Measures of Well-Being." *American Psychologist* 3: 117-124.

Campbell, Angus, Philip E. Converse, and Willard L. Rodgers. 1976. *The Quality of American Life: Perceptions, Evaluations, and Satisfactions*. New York: Russell Sage Foundation.

Cohn, R. M. 1977. Review of *The Quality of American Life*. *Contemporary Sociology* 6: 489-490.

Forti, Theresa Josephine. 1979. *Effect of Organizational Change on Well-Being Indices of a Louisiana Marianite Community in Response to Consultation*. New Orleans: Tulane University School of Public Health and Tropical Medicine (doctoral dissertation).

Glock, C. Y. 1976. "The Sense of Well-Being: Developing Measures." *Science* 194: 52-54.

Hall, John. 1976. "Subjective Measures of Quality of Life in Britain, 1971 to 1975: Some Developments and Trends." *Social Trends*, no. 7. London: Her Majesty's Stationary Office.

Headey, Bruce. 1980. "The Quality of Life in Australia." *Social Indicators Research*, in press.

Land, Kenneth C. 1978. "Developing Methods for Measuring Well-Being." *Contemporary Sociology* 7: 389-391.

Mason, Robert. 1978. Review of *Social Indicators of Well-Being*. *Social Indicators Research* 5: 369-376.

McKennell, Aubrey C. 1978. "Cognition and Affect in Perceptions of Well-Being." *Social Indicators Research* 5: 389-426.

McKennell, Aubrey C., and Frank M. Andrews. 1980. "Models for Cognition and Affect in Perceptions of Well-Being." *Social Indicators Research* 8: 257-298.

McKennell, Aubrey C., Tom Atkinson, and Frank M. Andrews. 1980. "Structural Constancies in Surveys of Perceived Well-Being." In Alexander Szalai and Frank M. Andrews, eds., *Comparative Studies on the Quality of Life*. London: Sage.

Michaelsen, Larry K., Donald Murray, Neil J. Dikeman, Howard Vanauken, and Marjory Earley. 1976. "The Quality of Life in Oklahoma, 1976." Norman, Okla.: Center for Economic and Management Research, College of Business Administration, University of Oklahoma.

Michaelsen, Larry K., Neil J. Dikeman, George W. England, Renee Alonso, Marjory Early, and Ellen Harrington. 1980. "The Quality of Life in Okla-

homa, 1979." Norman, Okla.: Center for Economic and Management Research, College of Business Administration, University of Oklahoma.

Phillips, Derek L. 1978. "What's One to Believe?" *Contemporary Sociology* 7: 392-395.

Rodgers, Willard L. 1977. "Work Status and the Quality of Life." *Social Indicators Research* 4: 267-288.

Rodgers, Willard L., and Philip E. Converse. 1976. "Measures of the Perceived Overall Quality of Life." *Social Indicators Research* 2: 127-152.

Sontag, Suzanne, Margaret M. Bubolz, and Ann C. Slocum. 1979. *Perceived Quality of Life of Oakland County Families: A Preliminary Report.* East Lansing, Michigan: Michigan State University Agricultural Experiment Station.

Vaughan, Denton R., and Clarise G. Lancaster. 1979. "Income Levels and Their Impact on Two Subjective Measures of Well-Being: Some Early Speculations from Work in Progress." *Proceedings of the American Statistical Association.* Washington, D.C.: American Statistical Association.

Wilkening, E. A., and D. McGranahan. 1978. "Correlates of Subjective Well-Being in Northern Wisconsin." *Social Indicators Research* 5: 211-234.

Zautra, Alex, Ernest Beier, and Lawrence Coppel. 1976. "The Dimensions of Life Quality in a Community." *American Journal of Community Psychology,* December.

Response to Comments

Allen R. Wilcox
University of Nevada, Reno

The variety of issues raised by Andrews's comments could easily lead to a lengthy and reverberating dialog. In the context of this particular exchange of ideas, however, I will resist the temptation to be expansive and limit myself to a small set of comments on comments.

ON FRACTIOUSNESS AND DOING BETTER

Andrews takes note of the fractious tone of my presentation. To the extent that any fractiousness did slip into my tone, I hope its effect was to stimulate dialog, not to give personal offense or to appear to slander pioneering social science research. Any innovative work will and should engender critical commentary as well as a certain amount of apotheosizing. Old classics such as *The American Voter* and newer ones such as *Participation in America* (see Rusk, 1976, and Wilcox, 1980) have and will continue to undergo careful scrutiny.

This process can take several forms. Andrews appears to favor a confrontation of empirically oriented studies since he refers in one place to the need to demonstrate that something better can be done and in another to the necessity of presenting significant empirical evidence. These are indeed indispensable activities, ones in which I expect myself, the authors of SIWB and QOAL, and many others to participate. However, they are not activities indispensable in every context, including that of my original essay and the present exchange of views. I will consequently proceed with a few more observations unbuttressed for now by empirical evidence.

ON A CONSOLIDATED ASSESSMENT DIMENSION

I find Andrews's comments on a consolidated assessment dimension useful. I assume from his discussion of the size of automobiles that one appropriate model for uncovering such a dimension would be hierarchical factor analysis with the highest order factor representing a consolidated dimension. This

relatively inductive approach has a rich history in multivariate psychological research, and I see no reason why it cannot be applied to the subject matter of SIWB and QOAL (extensions of this approach are explicated in Cattell, 1979, and Royce, 1980).

It might also be possible, however, to follow a somewhat more deductive path, which leads into perhaps the major dissatisfaction I experienced with my own expressions of dissatisfaction. Although I remain quite skeptical about the use of satisfaction questions as indicators of perceived well-being or quality of life, it seemed to me then, as it does now, that the concept of satisfaction might be a pivotal one in research on subjective social indicators.

In my essay, I treated satisfaction as one among many types of emotional experience. It might instead be treated as a more abstract concept, defined within a formal system of concepts and propositions, and then brought down the ladder of abstraction to the level of empirical measurement. For example, the concept of satisfaction plays such a role in the conceptual edifice constructed by Ackoff and Emery in *On Purposeful Systems* (1972) wherein "an individual's degree of satisfaction with an object, event, property or properties of either, or a state, X, is his degree of intention to produce a nonchange in X" (1972, p. 101). It might be worth quoting further from Ackoff and Emery to indicate how they would bring their concept of satisfaction one step closer to empirical measurement:

> For example, if an individual is in a particular environment, S, and he is presented with two exclusive and exhaustive classes of courses of action—members of one will change the environment and members of the other will not—and the other conditions of an intention environment are met, then the probability that he will select a course of action that will not change S is his degree of satisfaction with S. The probability that he will select the course of action that will change the environment is his degree of dissatisfaction with S. If the first probability is greater than the second, he is said to be *satisfied*. If the second is greater, he is *dissatisfied*. If these are equal, he is *indifferent* to the situation and can be said to have no feelings about it (1972, p. 101).

The concept of satisfaction might also be linked to existing theories that do not refer to it directly. For example, I believe a case could be made for relating satisfaction to the "general incongruity adaptation level hypothesis" advanced by Struefert and Struefert (1978).[1] In both these examples, however, I believe careful examination of the theoretical exposition would cast doubt on the validity of operationalizations utilizing survey questions on satisfaction.

---

[1] The General Incongruity Adaptation Level (GIAL) postulated by Streufert and Streufert (1978) "would motivate cognitive activity whenever the general incongruity currently being experienced by the organism departs from the expected value, i.e., whenever inconsistency is experienced" (italics omitted). Within this formulation, dissatisfaction might be equated with the experience of inconsistency (perhaps with an added tinge of affect) and, conversely, satisfaction with the experience of consistency.

ON POSITIVITY BIAS

Andrews is particularly sensitive about my assertion that "the net effect of some of the distortions . . . is to register levels of satisfaction far above what they actually are." In retrospect, I probably should have placed a standard academic qualifier such as *probably* in that sentence to indicate the lack of empirical substantiation. However, I still believe the arguments I made regarding positivity bias (and lack of self-knowledge, which was also mentioned within the ellipses of the quotation) are quite plausible.

I would grant that under ordinary (i.e., noncrisis or nontraumatic) conditions most people may tend to adopt various adjustment strategies that lead them to register on the satisfaction rather than the dissatisfaction side of the interview ledger. (Inglehart, 1977, discusses this process quite concisely, particularly as it applies to variation across stable social categories.) But this is not to say that the ledger is an accurate record. A complete argument would draw on diverse theoretical perspectives and bodies of evidence, but I limit myself here to a brief elaboration of one argument from the original essay.

Let us assume, in addition to conscious psychological processes that find their way onto the survey ledger, the existence of unconscious processes. Experimental as well as clinical research appears to register strong support for this supposition (see Shevrin and Dickman, 1980). Further, the concept of repression has been an integral part of theorizing about the unconscious. Assuming repression does occur, it follows from such motivational theorizing that sentiments relating to dissatisfaction, not to satisfaction, would most likely be repressed. As Andrews notes, "there is an important psychological need to experience at least some positive elements in aspects of life that are closely linked to oneself," and I would add, conversely, not to experience (consciously at least) negative elements.

Estimating the extent of this phenomenon is clearly extremely difficult, and it may be that Andrews would view this as another item on a future research agenda. If so, present data on levels of satisfaction should be treated with extreme caution.

This topic must be taken one step further. Even if reported levels of satisfaction were assumed to be accurate, this would not imply *approval* of those levels, whatever they might be. Individuals may, for example, lower their aspirations in the face of a tyrannical political regime and express high levels of satisfaction. An outside observer, however, might hope for eventual dissatisfaction, judging satisfaction in such a case to be inversely related to quality of life. And this brings us back to the meaning of that latter phrase.

ON QUALITY OF LIFE AND ITS ASSESSMENT

Whatever we might take this highly elusive and abstract phrase to mean, its operationalization must clearly go far beyond self-reports in response to survey

questions. My own assessment of a series of possible claims that might be made for SIWB and QOAL was that they come "closest to supporting a claim of being useful exploratory work into perceptions of satisfaction, happiness, and other types or aspects of well-being." The quote from Rescher that Andrews notes was part of an attempt to indicate that such perceptions cannot be taken as accurate indications of quality of life (unless that phrase is quite narrowly construed), a point elaborated later in the essay. I believe Andrews is correct in the sense that the Rescher quotation does overstate the case for objective conditions and that, as Andrews states, "both SIWB and QOAL have some relevance to the topic of life quality." Although I made some tentative suggestions in the last section of the essay, I would admit that I have not adequately explicated what that relevance might be. Neither, I might venture, have the authors of SIWB and QOAL.

In response to my request for conceptual clarification, Andrews notes that "anyone who takes the time to examine the books can find out what was done and what was learned—and that is what was intended." This appears to amount to saying don't pay attention to what we said we were doing, pay attention to what we did. This approach has value in any critical appraisal of research (as well as judgments about political pronouncements). However, there is also a norm in social science that suggests that there should be a strong link between theoretical and conceptual discourse and empirical application. I think this norm is worth preserving.

The meaning of the term quality of life is a topic that should engage the energies of social scientists, philosophers, and, indeed, entire societies. Perhaps within the context of a more cooperative and complex enterprise, we can avoid becoming mired down in a millenial debate and somehow trudge forward. We can but hope that the seminal contributions made in SIWB and QOAL will come into clearer focus and be built upon with time.

REFERENCES

Ackoff, Russell L., and Fred E. Emery. 1972. *On Purposeful Systems.* Chicago: Aldine-Atherton.

Cattell, Raymond B. 1979. *Personality and Learning Theory, Volume I: The Structure of Personality in Its Environment.* New York: Springer.

Inglehart, Ronald. 1977. "Values, Objective Needs, and Subjective Satisfaction among Western Publics." *Comparative Political Studies* 9: 429-458.

Royce, Joseph R. 1980. "Personality as an Adaptive System." *Systems Science and Science.* Proceedings of the 24th Annual North American Meeting of the Society for General Systems Research, San Francisco, pp. 226-233.

Rusk, Jerrold G. 1976. "Political Participation in America: A Review Essay." *American Political Science Review.* 70: 583-591.

Shevrin, Howard, and Scott Dickman. 1980. "The Psychological Unconscious: A Necessary Assumption for All Psychological Theory?" *American Psychologist* 35: 421-434.

Streufert, Siegfried, and Susan C. Streufert. 1978. *Behavior in the Complex Environment.* New York: Wiley.

Wilcox, Allen R. 1980. "Need Theory and Political Participation: Toward An Agenda." Paper presented at the 1980 Annual Meeting of the International Society of Political Psychology, Boston, June 1980.

Surveys of Subjective Phenomena: A Working Paper

Charles F. Turner
National Research Council
National Academy of Sciences

OVERVIEW

In a broad sense, the present chapter is concerned with the interrelation of psychology and survey research. Our more particular concerns are prompted by anomalies in the results of surveys that purport to measure identical subjective phenomena in a comparable manner. The burden of these anomalies, we submit, is sufficiently weighty to motivate a reconsideration of the psychological assumptions underlying the practice of survey research.

In the present chapter we assemble a variety of new evidence and suggest some tentative hypotheses concerning the types of subjective phenomena particularly vulnerable to artifacts of measurement. The present article is a montage; it presents abbreviated analyses of a number of examples rather than intensive case studies. The future, we hope, will provide the time and resources for a careful reconsideration of each of these examples.

AREA OF INQUIRY

Traditionally the term *subjective* has been used to denote those phenomena which are, in principle, directly observable only by subjects themselves. Phenomena of this sort include those commonly labeled "attitudes," "beliefs," and "opinions." These may be conceptually distinguished from other phenomena that although frequently measured by subjective means (i.e., self-report) are theoretically amenable to independent confirmation. In accord with traditional usage, we treat the possibility of independent verification (corroboration) as a litmus test for classifying phenomena as subjective or nonsubjective. This position assumes the existence of a nonsubjective (i.e., objective) reality whose

This working paper is based on a talk delivered at the 86th annual convention of the American Psychological Association in Toronto, Canada, August 28-September 1, 1978. The views expressed herein are the sole responsibility of the author; they should not be attributed to the National Academy of Sciences or the National Research Council.

properties are potentially discoverable through some consensual process. (One, of course, need not make such an assumption; see, e.g., the writings of Bishop Berkeley or the radical skepticism of Descartes' first meditation.)

So, for example, while we may measure age by asking respondents to report it, it would be theoretically possible to obtain independent evidence from other witnesses. For this reason we would not label chronological age per se as a subjective phenomenon. In theory, many other phenomena may also be measured independently of a subject's own report (e.g., educational attainment, geographic mobility, fertility history, family structure, income). However, many important phenomena are inherently subjective and thus immune to third-party verification. In particular, we have no direct knowledge[1] of an individual's attitudes, beliefs, or opinions. The present inquiry focuses upon survey measurements of such *subjective* phenomena.

IMPORTANCE OF AREA

Traditionally, national statistics have been the domain of demographers and economists. Inquiries made by the U.S. Bureau of the Census have generally been limited to assessments of the size and distribution of the population and a variety of other phenomena that are at least theoretically amenable to third-party[2] corroboration, for example, age, income, educational attainment. This does not mean, of course, that subjectivity does not contaminate such assessments; the use of self-report inevitably raises this issue. However, the validity and reliability of survey estimates of such phenomena may be supported by studies using independent sources (e.g., birth and earnings records) to estimate the magnitude and sources of error introduced by the exclusive use of information supplied by subjects themselves (see, e.g., the work of the evaluation and research program of the Census Bureau).

Our concern centers upon several measurements that have been used in recent years as subjective social indicators. It would, however, be a gross oversimplification to distinguish between the "old objective" and the "new subjective" indicators. Some well-known measurements, such as the national unemployment statistics, contain fundamentally subjective elements. For example, to be classi-

---

[1] Observation of a subject's behavior can only provide an indirect measure of a subject's attitudes, beliefs, etc. Conceptually, one can distinguish between behaviors and attitudes; in practice, experience affirms that an individual's behavior need not accord with beliefs. Indeed, one should bear in mind that typical measurements of subjective states involve inferences made from a subject's behavior(s), i.e., verbal behavior in response to a question. Such response behaviors are used to infer the state and structure of subjective reality. However, as behaviors, such responses are only *indirectly and assumptively* linked with the inferred phenomenon (the subjective state of the individual). Among others, misreporting and measurement artifacts are common pitfalls in this inference process. The general topic of attitude behavior inconsistency has, itself, received considerable attention in the social psychological literature (see, e.g., the recent review by Schuman and Johnson, 1976).

[2] In this regard, we note the traditionally unproblematic concept of a *true population value* in studies of sampling and response error in census work. This concept becomes more difficult to sustain (at least intuitively) when discussing the measurement of subjective phenomena (cf. Waksberg, 1975).

fied as unemployed in the U.S. Bureau of Labor Statistics (BLS) index, workers must be both out of work and looking for work. While the definition and interpretation of the criterion, looking for work, are not left solely to the discretion of respondents, statements that a worker has "checked with friends or relatives" during the last 4 weeks satisfy the second criterion. This component of the unemployment index is fundamentally subjective because the meaning and interpretation of the statement, "checked with friends . . . ," is supplied by the respondents and is inherently unverifiable. Similarly, evaluations of the National Crime Surveys by the U.S. Bureau of the Census (cf. Cowan et al., 1978; Gibson et al., 1978) have pointed up the subjective components of crime victimization statistics derived from subjects' self-reports. In the latter instance, the meaning and interpretation of phrases such as "Did anyone try to attack you . . . ?" are supplied by the respondents and may differ from respondent to respondent and across the contexts in which the measurements are made (cf. Martin, 1978, 1981).

In recent years, national statistics have come to include an important and rapidly growing complement of statistics designed to measure subjective phenomena. For example, the *Social Indicators* (U.S. Department of Commerce, 1973, 1977) program begun by the Office of Management and Budget incorporates measurements of a wide range of subejctive phenomena. (An independent review (Caplan and Barton, 1976) of the use of the first *Social Indicators* volume concluded that there was a need for Federal statistical compendia to "go beyond objective indicators and provide subjective measures of life experience and social well-being.") The most recent volume of the *Social Indicators* series argues that such measures provide a vitally needed supplement to traditional national statistics,

> The basic reason for including such subjective measures in this report despite the difficulties in their interpretation is that they offer a vital dimension in developing a comprehensive description of the condition of our society and the well-being of its members. The bulk of the information presented relates to people's objective situation or condition—their jobs, their income, their health status, etc. The main purpose of the attitudinal measures is to provide some insight as to how people perceive certain aspects of these conditions. Such data are an essential source of information on people's values and aspirations. (U.S. Department of Commerce, 1977, p. xxvi)

For similar reasons, the National Science Board's recent series of reports (1973, 1975, 1977) on the state of science in the United States has incorporated a concluding chapter on public attitudes toward science and technology. Interest in this topic follows from the fact that financial support, the imposition of legal constraints (e.g., regulation of recombinant DNA research), and the recruitment of young people into the scientific professions depends, in part, upon public perceptions of science. In the Board's own words,

> Public attitudes affect science and technology in many ways. Public opinion sets the general environment and climate for scientific research and technological development. It is influential in determining the broad directions of research and innovation, and through the political process, the allocation of resources for these activities. In addition, public attitudes toward scientists and engineers and their efforts affect the career choices of the young by influencing their decision to enter these fields. (National Science Board, 1975, p. 145)

The increasing importance of measures of subjective phenomena in Federal statistical programs is paralleled by a growing range of relevant research activities[3] in the academic community. This work has included psychological studies of well-being (e.g., Campbell et al., 1976; Andrews and Withey, 1976; Bradburn, 1969; Staines and Quinn, 1979), investigations by sociologists of trends across time in sex-role stereotyping and the tolerance of nonconformity (e.g., Davis, 1975a; Duncan, 1979; Mason et al., 1976), and work by economists on the relationship of economic development to individual happiness (e.g., Easterlin, 1974).

Similar trends are evident around the world. Since 1970, Britain has issued annual reports entitled *Social Trends*. These reports pay considerable attention to measures of subjective phenomena. Indeed, they argue that:

> The more one considers [indicators of the quality of life], the more one is persuaded that the way forward lies not in adding more measures of conventional hard statistics, but rather in supplementing the existing ones by adding ... a dimension of the satisfaction (happiness, contentment, psychological well-being, etc.) felt by those who constitute the community and are the final consumers of society's "goods" and "bads" and therefore the best judges of society's performance. In short it is the very thoroughness [of earlier work on hard statistics] that compels one to turn to subjective social indicators and to the problems of reliable quantification of states of mind and mood.... (Abrams, 1973, p. 36)

Moreover, a recent United Nations report, *Toward a System of Social and Demographic Statistics* (1975), echoes this concern for the inclusion of measures of subjective phenomena in national statistics. In a chapter beginning,

---

[3]The emergence of such efforts is reflected by a number of occurrences. For example, the social science community has established a national data program with the specific aim of providing data from representative national samples to enable the construction of long term time series of both demographic and attitudinal indicators. Similarly, there was a unique, albeit short-lived, attempt to provide Federal policymakers with easy access to an ongoing series of national surveys. The latter project was funded by the RANN division of the National Science Foundation (NSF) for the use of the following agencies: Department of Agriculture; Department of Health, Education, and Welfare; Department of Housing and Urban Development; Department of Transportation; Office of Management and Budget; President's Commission on Gambling; and NSF. For a history of this pioneering but ill-fated program, see Rich (1975).

"The need for social indicators . . .," the report observes that

> In dealing with social questions, however, we may also be interested in subjective information relating to how much people in general know about an issue, how much importance they attach to it and what kind of solution they think would be desirable.
>
> Public opinion surveys provide a means of obtaining some light on such matters. It would be interesting to know, for instance, what issues are commonly regarded as major problems and how the ranking of these issues changes with time. It would also be interesting to know how far the public connects one issue with another: does it believe, rightly or wrongly, that the great increase in pollutants in recent years is associated with activities and processes that contribute to the rising standard of living; does it believe, rightly or wrongly, that the scale and organization of modern enterprise, which also contribute to the standard of living, are associated with industrial unrest and alienation?

COMPARABILITY OF SUBJECTIVE SOCIAL INDICATORS

Survey measurements of subjective phenomena are made by many organizations. In the United States, nonfederal sources produced the majority of the subjective social indicator measurements reported in *Social Indicators; 1976.*

The use of nongovernmental surveys to collect such attitudinal information reflects both ideological and practical considerations. It is sometimes argued that on principle governments ought not to ask their citizens for certain information even though it might be useful in developing government policy. Concerning religious beliefs, for example, it has been argued that "no federal statistical agency has any business whatever inquiring about anyone's religious beliefs, even though information about the distribution of beliefs in the population is pertinent to various federal policies" (Duncan, 1972, p. 152).

At a practical level, doubts have been raised about the ability of governments to obtain reliable information on sensitive topics, for example, antigovernment sentiments.[4] Thus, in an appendix to the report of the Presidential Commission on Federal Statistics, Sheldon (1971) proposed a division of labor between government and nongovernment statistical organizations,

> The development of time series information covering subjective dimensions as well as topics presumed to be politically sensitive will continue to

[4] A recent survey conducted under the auspices of the Committee on National Statistics (National Academy of Sciences) suggests that such concerns may be overstated. In a survey conducted jointly by the Census Bureau and the Survey Research Center of the University of Michigan, virtually identical results were obtained by both organizations on a large variety of items, including several measuring antigovernment attitudes. The only significant differences occurred for items specifically asking about social surveys and survey-taking organizations (see Goldfield et al., 1977; National Academy of Sciences, 1979).

> be the primary responsibility of non-governmental research and university centers.... This work requires considerable conceptual innovation and field experimentation activities particularly appropriate to institutions independent of governmental agencies. (p. 421)

Use of data from a variety of sources, however, inevitably raises questions of comparability. Despite one's hopes, comparability of measurement does not occur naturally (cf. Hunter, 1977; Ho et al., 1974); experience indicates that it is, rather, the result of careful standardization of research procedures and the continuous monitoring of performance. For example, attempts by analytical chemists to achieve comparability of measurement across laboratories involved a long history of standardization of research practices and the development of methods for collaborative tests *across laboratories* (cf. Youden and Steiner, 1975).

We believe that emerging evidence of noncomparability in survey measurements of subjective phenomena argues both for a consideration of appropriate techniques for assuring data comparability when measurements are derived from several sources and a reconsideration of the psychological assumptions that underlie the practice of survey research.

THE PROBLEM

The use of replicated time series of subjective social indicators has given birth to some disagreeable progeny. Most irritating has been the multiplication of instances in which allegedly identical measurements of subjective phenomena have differed both substantially and significantly between surveys (cf. Turner and Krauss, 1978; Martin, 1981); in some instances, discrepancies of 15 percentage points in the univariate distributions have been observed. These discrepancies prompt a number of questions; one would like to know for example:

Why these measurements disagree
Whether these disagreements are symptomatic of a larger problem or whether they are restricted to a few isolated cases
Whether there are any organizing principles that could provide a typology of those indicators more (and less) likely to give discrepant results

In the following pages we review examples drawn from a variety of social indicator projects. These examples shed some light upon these issues and illustrate the need for further research. In addition, we test some initial organizing principles concerning the types of subjective indicators that are particularly vulnerable to artifacts of measurement. We do so not in the hope of elucidating final principles but rather to provide initial hypotheses around which research may be organized. As a starting point, we take the recent suggestion (Turner and Krauss, 1978, p. 468) that discrepancies in estimates of subjective social indicators may be a function of the survey questions, them-

selves, and the phenomena they intend to measure. In particular, it was hypothesized that the discrepancies would be concentrated among those indicators that involved survey questions that

1. Were most amorphous in their meaning, for example, those seeking to assess "confidence," "trust"
2. Were most ambiguous in their referents, for example, those inquiring about the "people running organized religion"
3. Involved the most arbitrariness in the selection of a response category, for example, great deal versus some confidence

Using these hypotheses as our point of departure, let us turn to our first example.

THE CASE OF HAPPINESS

This example is the first of several where surveys disagree. In the fall of 1977, we began an investigation of responses to national survey questions on personal happiness. These questions have been incorporated in research attempting to define the nature of social well-being and to produce sensitive estimates of life satisfaction (e.g., Gurin et al., 1960; Bradburn, 1969; Campbell et al., 1976). Much of this work has gone beyond the notion that responses to a single question are ideal measures of subjective well-being. Nonetheless, responses to the simple question:

> Taken all together how would you say things are these days—would you say that you are very happy, pretty happy, or not too happy?

have been tracked from the year 1957 (cf. Gurin et al., 1960), and both the trends across time and differences between nations in response to this question have been discussed by several authors (e.g., Easterlin, 1974; Davis, 1975b; Campbell et al., 1976; Andrews and Withey, 1976). Moreover, the responses to this happiness question have been used as a validity criterion in the development of more elaborate indices of life satisfaction. Thus, responses to this question are of substantial importance and interest in their own right.

Figure 1 presents two independent series of happiness estimates derived from surveys conducted by the Survey Research Center of the University of Michigan (SRC) and the National Opinion Research Center of the University of Chicago (NORC). This figure shows that there are not only discrepancies in estimates of the absolute *levels* of happiness but also that the trends in the two series appear to diverge in direction. One series shows an apparent increase while the other series registers a decline in happiness.

We first noted this disagreeable result in the fall of 1977, and it was the subject of preliminary discussions with an ad hoc working group, which met to discuss

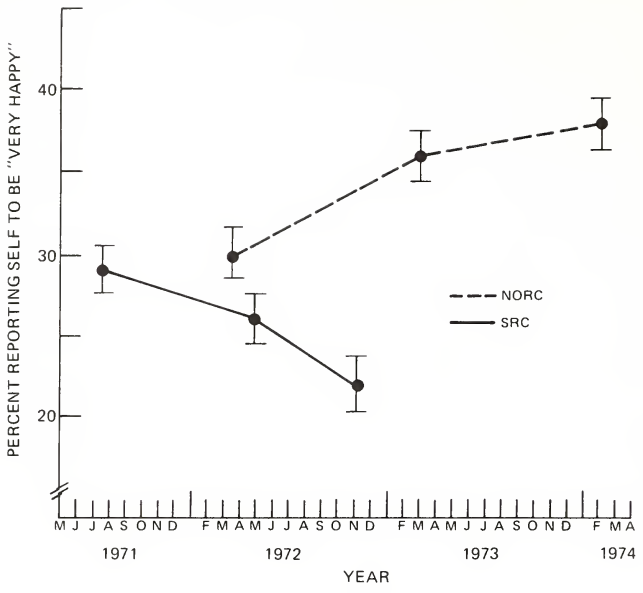


Figure 1. Trends in self-reported happiness, 1971-1973. Estimates are derived from sample surveys of noninstitutionalized population of the continental United States, aged 18 and over. Error bars demark ±1 standard error around sample estimate. *Source:* NORC, *National Data Program for the Social Sciences: Codebook, 1972-1974.* SRC, estimates from Campbell et al. (1976); survey dates from Campbell et al. (1976), Andrews and Withey (1976), and J. Varva (personal communication).

the discrepancies that had been observed in the "confidence in institutions" series.[5] Subsequent examination of the two happiness series has caused us to doubt the validity of the comparison shown in figure 1. Examination of the questionnaires used by NORC and SRC reveals some differences in question wording:

> Taken all together, how would you say things are these days—would you say that you are very happy, pretty happy, or not too happy? (NORC)
>
> Taking all things together, how would you say things are these days—would you say you're very happy, pretty happy, or not too happy these days? (SRC)

The SRC version of the happiness question repeats "these days" at the end of the question, whereas the NORC version does not. Thus, it might be argued that these questions were indicators of slightly different phenomena, although

[5]The meeting was hosted by the Institute for Research in the Social Sciences, University of North Carolina, and attended by J. Davis, O. D. Duncan, E. Martin, F. Munger, R. Parke, H. Schuman, T. Smith, G. Taylor, and C. Turner. We first presented these data at the meetings of this working group. Subsequently, one member of the group conducted his own study of these data. He independently reached a conclusion similar to ours concerning the likely cause of the observed discrepancies (cf. T. Smith, 1979).

admittedly one might expect the trends across time to be parallel rather than divergent. However, it could be that the divergences are more apparent than real. For example, national happiness might have fluctuated rapidly between 1971 and 1973, and thus, these data could be reliably mirroring *month-to-month* changes taking place in the national population.

While such arguments can be made, they are apologies rather than explanations. The NORC and SRC data have been treated as a *unitary* time series by several authors, despite differences in wording (e.g., Campbell et al., 1976, p. 26; Andrews and Withey, 1976). Moreover, large month-to-month fluctuations in these indicators would preclude use of annual and biennial reporting schedules, which have been the rule in the social indicators field. (If the trends are month-to-month, then yearly or biennial measurements (taken in different months) may prove uninformative or misleading.)

Thus, despite their limitations, the data shown in figure 1 prompted us to speculate about the causes of the observed discrepancies. Our initial hypothesis followed from the fact that NORC altered its questionnaire in 1973 so that a question about marital happiness,

> Taking things all together, how would you describe your marriage? Would you say that your marriage is very happy, pretty happy, or not too happy?

immediately preceded the general happiness question.

We hypothesized that insertion of this marital happiness question created an artifactual response bias. Our initial examination of this hypothesis (tables 1 and 2) indicated that:

1. There was a high correlation between responses to the marital and general happiness questions
2. The marital happiness question elicited a relatively high proportion (.6) of very happy responses
3. The increase in overall happiness between 1972 and 1973-1974 in the NORC series occurred only among married persons.

This last finding[6] was particularly important because the hypothesized context effect could only have occurred for married individuals. Unmarried persons were not asked, of course, about the happiness of their marriages.

While any comparison of the NORC and SRC happiness series admits to a plethora of alternate explanations (e.g., wording effects, house effects, short term temporal variations), the results of our initial explorations encouraged us

[6] The divergence in these trends is only suggestive since analysis of the data indicates that a model positing only effects for year and marital status provides a tolerable fit for these data. An interaction term for the context effect (year X marital status) is not strictly required.

Table 1. Association Between Responses to Marital Happiness and General Happiness Questions: 1973-1977

Marital happiness	General happiness (percent)			N^a
	Not too happy	Pretty happy	Very happy	
Not too happy	65	32	3	150
Pretty happy	11	78	11	1,502
Very happy	5	38	57	3,408

Note: $\chi^2 = 1{,}094.6$; df = 4; $p < .0001$; $\gamma = +0.75$

[a]Chi-square statistics were adjusted for design effects of NORC's clustered sample design by using a deflated sample size ($N' = 0.66N$) in computations. Analysis of the intracluster correlations (median 1973-1978 $r_i = 0.02$) for the happiness item indicates that this correction provides a conservative estimate of the relevant sampling errors. Sample sizes shown are raw figures; they do not reflect deflation or weighting of sample by number of eligible adults in household.

to seek a better test for our hypothesis. We were furtunate to discover a wealth of information on happiness collected during this period. Between April 1973 and May 1974, the National Opinion Research Center with the support of the RANN division of the National Science Foundation conducted a series of pilot surveys to provide continuous monitoring of public opinion for policymakers in eight Federal agencies. At intervals of approximately 1 month, NORC drew samples of the national population for interview. While the content of the surveys varied from month to month, the happiness item was included in every cycle of NORC's Continuous National Survey (CNS).

Table 2. Married and Unmarried Respondents Reporting Themselves To Be Very Happy: 1972-1974

Sample	1972 (percent)	1973 (percent)	1974 (percent)	χ^2 for temporal change[a]	
				1972 vs. 1973 vs. 1974[b]	1972 vs. 1973 + 1974[c]
Married	33.5	42.7	44.6	22.0 ($p < .001$)	21.5 ($p < .001$)
Not married	17.9	20.0	19.6	0.4 (NS)	0.4 (NS)
Total	29.7	36.8	38.4	20.1 ($p < .001$)	19.6 ($p < .001$)

NS Not significant.

[a]Chi-square statistics were adjusted for design effects of NORC's clustered sample design by using a deflated sample size ($N' = 0.66N$) in computations. Analysis of the intracluster correlations (median 1973-1978 $r_i = 0.02$) for the happiness item indicates that this correction provides a conservative estimate of the relevant sampling errors. Sample sizes shown are raw figures; they do not reflect deflation or weighting of sample by number of eligible adults in household.

[b]df = 2.

[c]df = 1.

These data allow us to compare responses across time for two identically worded questions in surveys conducted by the same research organization. This provides a control for both wording differences and any possible organizational idiosyncracies (e.g., variations in interviewer training). We obtained a copy of these data during February 1978 and set to work examining the plausibility of our context hypothesis for the misbehavior of the happiness time series.

Figure 2 presents a graphic summary of our findings. Specifically, we found that for unmarried individuals, yearly estimates derived from the NORC General Social Survey (GSS) and the monthly estimates from the CNS were in general agreement. This is not to say that the estimates were identical. However, observed discrepancies were well within the range expected on the basis of sampling error. In short, unmarried men and women responded to the happiness question in the same manner in the 1972-1974 GSS and the 12 cycles of the CNS.

For the married respondents, a rather different result emerged. In particular, the GSS happiness estimates exhibit a sharp rise ($\chi^2 = 21.5$; df = 1; $p < .0001$) between 1972 and 1973-1974 (change = +10 percent), but the 12 monthly

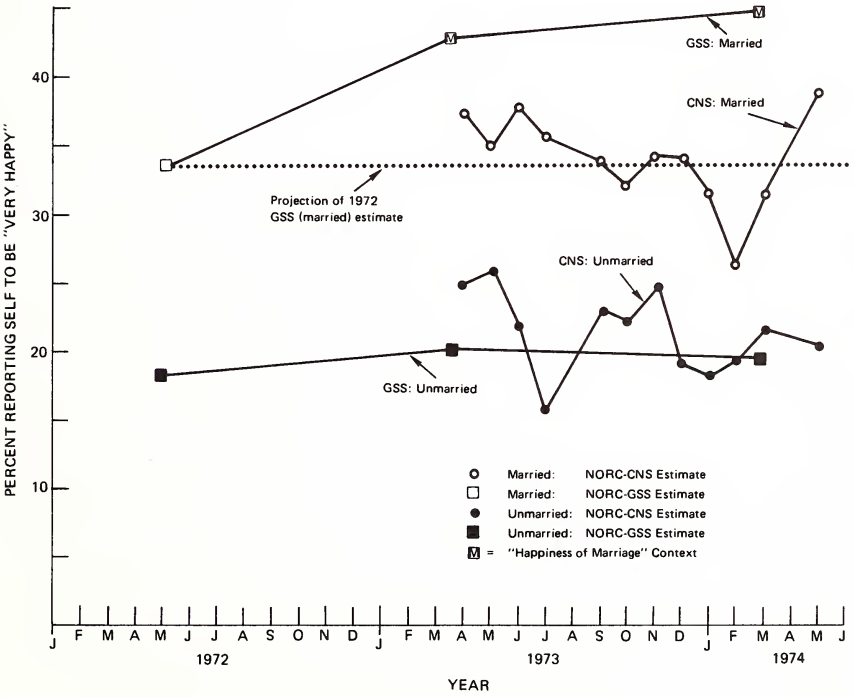


Figure 2. Variations in response to NORC happiness question for married and unmarried respondents in the General Social Surveys (GSS) and Continuous National Surveys (CNS). Estimates are derived from samples of approximately 1,000 (GSS) and 440 (CNS) married respondents and 500 (GSS) and 220 (CNS) unmarried respondents.

estimates derived from the CNS evidence no similar trend. Moreover, the CNS happiness estimates are consitently below those of the GSS (average difference = 9 percent). Indeed, as figure 2 shows, the 1972 GSS measurement—when the marriage question was not included—provided a better prediction of the CNS estimates in 1973 and 1974 than the actual GSS estimates in those years.

Although other hypotheses might be supported, we submit that (1) the internal evidence of a temporal trend *only* for married GSS respondents and, (2) the predictability of the 1973-1974 CNS data from 1972 GSS data provide strong support for the hypothesis that a response bias arose from the insertion of the marital happiness question in the 1973 and 1974 General Social Surveys.

The implications of this artifact are substantial. The GSS provides a major source of data for the social science community, and it was the largest source of subjective social indicator data for the Federal compendium, *Social Indicators, 1976* (U.S. Department of Commerce, 1977). It is not unreasonable to anticipate publication of substantive interpretations for the post-1972 rise in GSS happiness series (e.g., as an effect of the end of the Vietnam war). Such interpretations, of course, would be misleading. This increase in national happiness appears to result from changes in the content of our surveys and not from changes in the subjective state of the national population.

What do we learn from this example? Recalling the general hypotheses outlined in the introduction, we observe that the

1. Concept of happiness is notably amorphous
2. Happiness question involved considerable arbitrariness in the choice of a response category, for example, what is the difference between being very happy and pretty happy?
3. Question may be one to which individuals do not give considerable thought—at least as formulated in this item (i.e., Am I happy?)

FERTILITY EXPECTATIONS

Lest the reader be misled by our first example, we hasten to note that we do not believe that all survey measurements of subjective phenomena are equally vulnerable to artifactual biases. Rather, we wish to delineate areas in which artifact-induced discrepancies might be expected and the factors likely to cause such misbehavior. With this purpose in mind, let us consider some alternative estimates of the fertility expectations of American women.

The U.S. Bureau of the Census conducts annual surveys of the birth expectations of American women. The data from such surveys are potentially useful in predicting fluctuations in the birth rate. (Longitudinal studies indicate that those measures do predict subsequent fertility; e.g., Wilson and Bumpass, 1973; Freedman et al., 1975; Goldberg et al., 1959.) Clearly, the phenomenon being measured in such surveys is highly subjective. The intention or expectation of

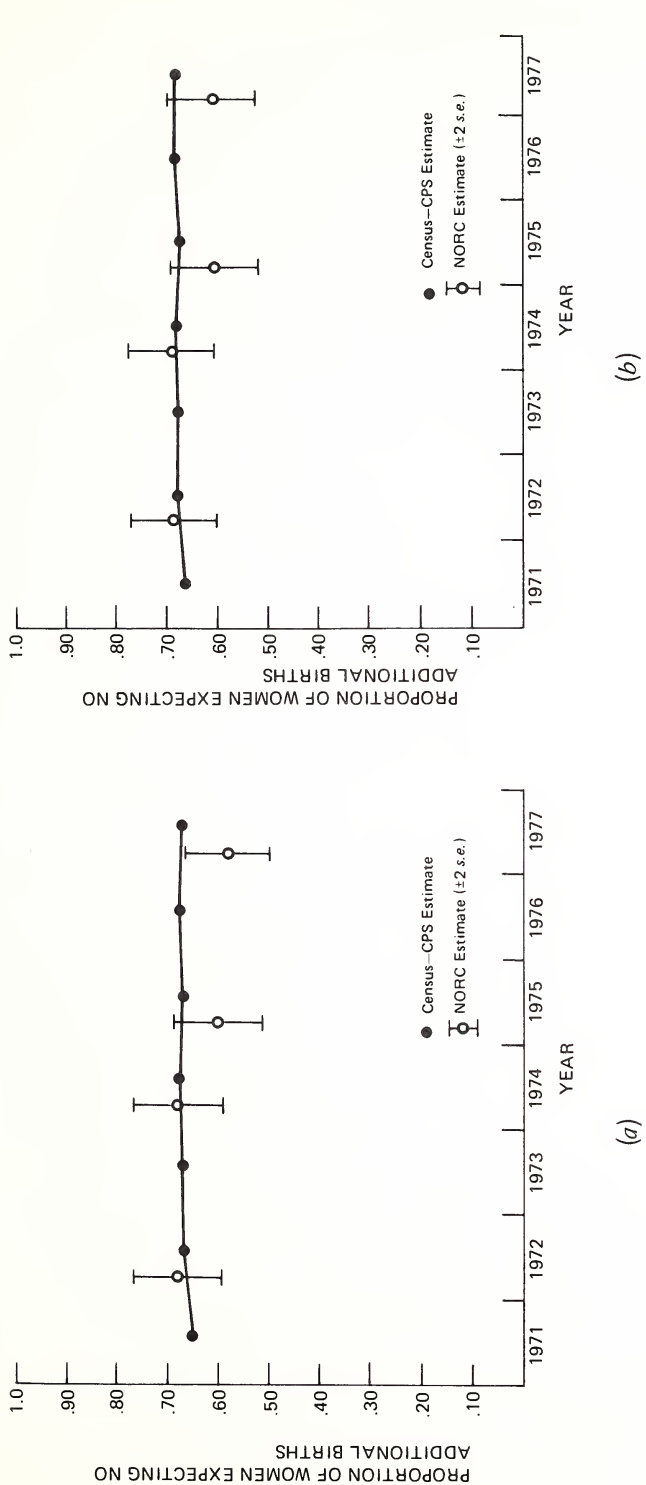


Figure 3. Estimates of fertility expectations of American women: proportion of women expecting no further children in (*a*) all future years and (*b*) next 5 years. Samples included only married women aged 18 to 39; sample sizes in each year were approximately 4,000 (Census-CPS) and 220 (NORC).

future pregnancy is not a datum subject to external verification. One must rely solely upon respondents' assessments of their own expectations or intentions.

Estimates from the Census Bureau's Current Population Survey (CPS) are shown in figure 3 together with estimates derived from a related question on birth expectations asked in the 1972 and 1975-1977 NORC GSS. It should be noted that the latter estimates are based on a very small sample; on the average, there were fewer than 250 married women aged 18-39 in the GSS samples. Thus, the standard errors for the GSS estimates are quite large (about 4 percent).

Comparing the two sets of data, we find that estimates of fertility expectations derived from NORC's GSS are quite consistent with those derived from the Census Bureau's CPS. In only one instance (of eight) does the NORC GSS estimate differ by more than two standard errors from the Census estimate.

The Lesson

The estimates of fertility expectations presented in figure 3*a* and 3*b* differed in their measurement in several ways. The content of the questionnaires used to derive the estimates varied, the organizations conducting the surveys were different, and even the wording of the questions varied. Table 3 presents the actual text of the questions used in each measurement. The questions asked about birth expectations varied slightly not only between the NORC and Census series but also within the Census series across time (i.e., 1972 vs. 1973-1977).

Because the comparison of these two series of birth expectations involved both wording and context differences and because the measurements were made at different times of the year by different organizations, the consistency of these estimates is particularly impressive. What lesson does this comparison teach us?

Table 3. Birth Expectations Questions Used by Census Bureau and NORC: 1972-1977

Organization	Years	Text
Census Bureau	1972-1974	1. Do you expect to have any more) children? 2. How many (more) do you expect to have? 3. How many (more) do you expect to have in the next five years?
	1975-1977	1. Looking ahead, do you expect to have any (more) children? 2. How many (more) do you expect to have? 3. How many (more) do you expect to have in the next five years?
NORC	1972-1977	1. Do you expect to have any (more) children? 2. How many (more)? 3. How many (more) in the next five years?

In terms of our initial hypotheses, we note:

1. The birth expectation question is relatively unambiguous in its meaning.
2. The response categories for the questions (e.g., 0, 1, 2, 3, . . . children) have a rather clear meaning.
3. The question deals with a topic to which most respondents (i.e., married women of childbearing age) should have given considerable thought, particularly because attitudes toward childbearing have behavioral consequences in the everyday life of the respondents, for example, contraceptive behaviors.

SCIENCE AND THE PUBLIC

Our next two examples of disagreement involve the measurement of public attitudes toward science and technology. These measurements were made in surveys commissioned by the National Science Board and conducted by the Opinion Research Corporation (ORC) of Princeton, N.J. The results of these surveys have been incorporated in the volumes *Science Indicators: 1972, Science Indicators: 1974,* and *Science Indicators: 1976* published by the National Science Foundation.

Our interest in these surveys was first aroused by an observation made during the analysis of the 1976 survey. In brief, the 1976 survey contained an anomaly that particularly concerned the staff person responsible for the chapter on public attitudes toward science. This anomaly had potentially destructive implications for national science policy and funding; thus, it is a most appropriate illustration of the dangers inherent in our inadequate understanding of the error structure of the data we employ as subjective social indicators.

The anomaly in the 1976 survey arose through an attempt to explore the meaning of public response to the following questions:

> Science and Technology can be directed toward solving problems in many different areas. In which of the areas listed on this card would you *most* like to have your taxes spent for science and technology? Please read me the numbers.
>
> (Card) 1. Reducing and controlling pollution. 2. Finding better birth control methods. 3. Weather control and prediction. 4. Space exploration. 5. Improving health care. 6. Developing/improving weapons for national defense. 7. Developing faster and safer public transportation for travel within and between cities. 8. Discovering new basic knowledge about man and nature. 9. Reducing crime. 10. Improving the safety of automobiles. 11. Finding new methods for preventing and treating drug addiction. 12. Improving education. 13. Developing/improving methods of producing food
>
> Please tell me the areas you would *least* like to have your taxes spent for science and technology. Again, please read me the numbers.

Data from the 1972 and 1974 Science Indicators surveys revealed that the public gave relatively strong endorsement to funding science in order to reduce crime (59 percent in 1972 and 58 percent in 1974), fight drug addiction (51 percent and 48 percent) and improve education (41 percent and 48 percent) and relatively weak support to science spending for such purposes as the development of faster and safer mass transportation (23 percent and 26 percent) and discovering new basic knowledge (19 percent and 21 percent). This ordering of public priorities contradicts many scientists' notions of where research could be useful, and it prompted an explicit study of this question in the 1976 survey.

In 1976, the Science Indicators survey was altered to incorporate the following questions *immediately prior* to the spending questions:

> This card lists a number of areas in which there are problems. In your view, in which of these areas could science and technology make a major contribution toward solving the problems? Please read me the numbers.
>
> (Card: same as above)
>
> In your view, in which areas could science and technology make little or no contribution? Please read me the numbers.

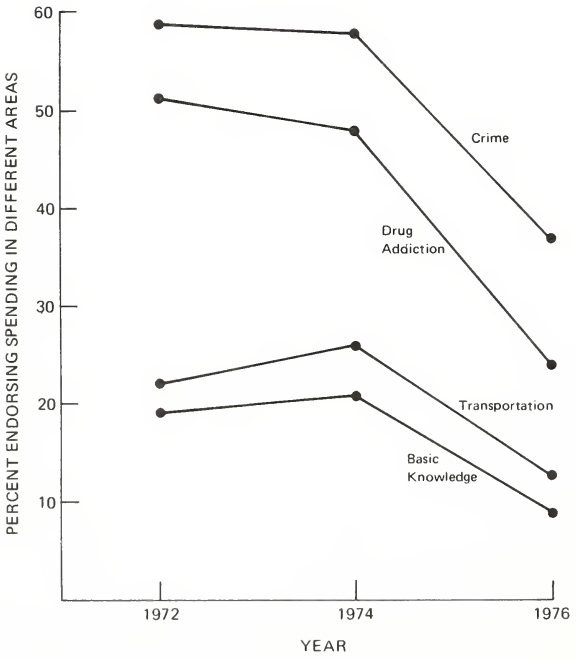


Figure 4. Endorsement of spending for science and technology in four areas. Sample size in each year was approximately 2,100. *Source:* Science Indicators surveys, 1972-1976.

Table 4. Endorsement of Spending for Science and Technology, Estimated by Science Indicator Surveys: 1972, 1974, and 1976

Area	Endorse spending (percent)			Change	
	1972	1974	1976	1972-1976	1974-1976
Improving health care	65	69	57	−8	−12
Reducing and controlling pollution	60	50	33	−27	−17
Reducing crime	59	58	37	−22	−21
Finding new methods for preventing and treating drug addiction	51	48	24	−27	−24
Improving education	41	48	33	−8	−15
Improving the safety of automobiles	38	29	15	−23	−14
Developing faster and safer public transportation for travel within and between cities	23	26	13	−10	−13
Finding better birth control methods	20	18	10	−10	−8
Discovering new basic knowledge about man and nature	19	21	9	−10	−12
Weather control and prediction	11	14	5	−6	−9
Space exploration	11	11	7	−4	−4
Developing or improving weapons for national defense	11	11	10	−1	−1
Average	34	34	21	−13	−13

Note: Estimates are for percent selecting areas in which they would most like to have taxes spent. See text for question wording.

Responses to the spending question in the 1976 survey were so unusual, however, that neither an analysis of the relationship between the perceived usefulness of science and the endorsement of spending, nor the spending time series itself appear in the final report of the National Science Board. Instead, a footnote observed that alterations in the ordering and content of the questions preceding the spending question preclude a valid comparison of the 1976 estimates to those obtained in previous years. This reticence is understandable.

Figure 4 and table 4 present the estimates derived from responses to the spending question in 1972, 1974, and 1976. These estimates show an apparently precipitous decline in public support of spending for science and technology. In two instances, this decline exceeded 20 percentage points.

This evidence of a massive drop in public support is, however, inconsistent with other independent evidence. The GSS has included an item on spending since 1973. The resultant NORC series show virtually constant levels of public support for science-related spending between 1973 and 1976 (table 5).

Because of the change in the questionnaire used in the Science Indicators surveys, the National Science Board chose not to present the data showing an apparent decline in public support for science spending. Instead, they argued in a footnote that

Table 5. Evaluation of Government Spending Programs, Estimated by NORC General Social Surveys: 1973-1976

Area	Spending too little or about right (percent)			Change	
	1973	1974	1976	1973-1976	1974-1976
Improving and protecting the Nation's health	95	95	95	0	0
Improving and protecting the environment	92	92	90	−2	−2
Halting the rising crime rate	95	95	92	−3	−3
Dealing with drug addiction	94	93	92	−2	−1
Improving the Nation's education system	91	91	90	−1	−1
Space exploration program	39	37	38	−1	+1
The military, armaments, and defense	60	67	71	+11	+4
Average	81	81	81	0	0

Note: Estimates are repercentaged to exclude "don't know" responses and no answers. Sample sizes in each year were approximately 1,500.

Question: We are faced with many problems in this country, none of which can be solved easily or inexpensively. I'm going to name some of these problems, and for each one I'd like you to tell me whether you think we're spending too much money on it, too little money, or about the right amount. First _____ are we spending too much, too little, or about the right amount on _____ ?

> The same [spending] question was used in the 1972 and 1974 surveys, but since it was not preceded in those years by the question about the *capabilities* of science and technology, the results are not strictly comparable to the 1976 results. (National Science Board, 1977, p. 180)

Our own analysis of the NORC spending data indirectly supports this argument. However, this position contradicts the prevailing wisdom among survey researchers. In their well-known book, *Response Effects in Surveys*, Sudman and Bradburn (1974) concluded that

> [the] position of a question [in the survey questionnaire] has by itself little biasing effect for behavioral items and a negligible effect for attitudinal items . . . [and] there do not appear to be any sizeable response effects associated with the placement of questions after related questions. (p. 33)

Context and Univariate Distributions

The National Science Board argued that their 1976 estimates of public endorsement of science spending were not comparable to those in previous years because of the different questionnaire contexts in which the spending question was embedded. Sudman and Bradburn argue that, in general, the effects of such context variations are negligible. Who is right?

In the absence of experimental evidence upon the specific context effects postulated by the National Science Board, it is difficult to assess the validity of their claim. The stability of public support for spending in other independent series circumstantially supports the Board's position. Similarly, recent experimental evidence (cf. Turner and Krauss, 1978, p. 466 ff.) indicates that context variations may produce fluctuations of up to 15 percentage points in the univariate distributions of some measurements.

While we cannot directly test the claim made by the Board, we have studied other circumstantial evidence on this question. This evidence arises because the Science Indicators surveys were amalgams consisting of several questionnaire sections sponsored by different organizations. The survey questions for the National Science Board's 1976 report were asked along with questions on hospitalization and medical expenditures, frequency of eating hamburgers, and the litter problem.

To control for context effects in these surveys, two different versions of the survey were administered. The different versions rotated the order of questionnaire sections in the questionnaire. Also, for some multipart questions, the sequence of individual parts of a question was varied. Each version of the questionnaire was administered to one-half of the sample.[7]

The first item in the Science Indicators section of the 1976 questionnaire assessed the prestige, or general standing, of various occupations including scientist and engineer. For the National Science Board, this question is of interest both because it serves as a surrogate measure of public attitudes toward

[7] Selection of respondents to receive form A or B took place as follows: The survey, itself, used three separately generated national samples. In sample 1, 60 sampling points received version A and 60 received B; selection was random. For sample 2, 30 sampling points received form A and 30 sampling points received form B, and for sample 3, 50 sampling points received form A and 50 received form B.

Randomization by sampling points (*N*s = 8) rather than individual respondents causes some difficulties. In particular, we know that clustering of sampling points results in a departure of sample efficiency from that of SRS designs; the extent in this instance is impossible to compute given the presently available data. In the analysis of the confidence items, Turner and Krauss (1978) found that the use of clusters of 15 as the unit for analysis produced a deflation of the effective sampling size from 1,500 to 1,000 (using the level of intracluster correlation as a deflator) (cf. H. Blalock, 1972, chapter 20; Kish, 1965).

Whether that result holds here is unknown; however, we would point out that anlaysis of the demographic characteristics of the samples (age, sex, education, income, number in household, religion, marital status) by form (A or B) revealed no significant differences (computations assume effective sample size to be deflated by 0.66 (i.e., *N* = 2,000 becomes *N* = 1,333)

Variable	χ^2	df	*p*
Sex	0.0	1	NS
Education	4.9	9	NS
Age	8.5	8	.40
Income	13.2	11	.27
Religion	4.6	4	.35
Marital status	0.8	4	NS

science and because the prestige of scientific occupations influences recruitment of talented young people into these professions. For social scientists, the responses to such questions are important because they provide the basis for well-known scalings of the socioeconomic status or prestige of occupations (Duncan, 1961; Hall and Jones, 1950; Treiman, 1977). These scales have been central to much recent work on social stratification (e.g., Blau and Duncan, 1967; Sewell and Hauser, 1975).

It is thus of considerable interest to know whether response to this question was affected by the context variation built into the Science Indicators survey. The survey question read:

> I am now going to read you a list of jobs and professions. For each one I mention, please choose the statement that best gives your own personal opinion of the prestige or general standing that such a job has.

The respondent was shown a card containing the responses: excellent, good, average, below average, and poor. Ratings were solicited for the 10 occupations shown in table 6.

The variation in the context and administration of the prestige question was twofold. First, the order of occupations listed in form A was reversed in form B (i.e., from businessman through accountant in form A, and in the reverse order in form B). Second, the placement of this question in the survey varied. In

Table 6. Variations in Excellent Ratings of Occupational Prestige, by Survey Form

Occupation	Excellent (percent)		Discrepancy (percent)	χ^2	p^a
	Form A	Form B			
Businessman	13.4	13.4	0.0	0.0	NS
Physician	47.6	56.4	−8.8	17.8	.005
Scientist	46.8	49.5	−2.7	8.4	NS
U.S. Representative in Congress	16.1	30.4	−14.3	54.1	.0001
Lawyer	24.0	38.7	−14.7	41.9	.0001
Architect	24.6	37.5	−12.9	29.9	.0001
Minister	39.0	38.3	+0.7	9.3	NS
Engineer	25.5	34.0	−8.5	18.9	.002
Banker	18.9	27.7	−8.8	26.7	.0001
Accountant for a large business	17.3	25.0	−7.7	28.9	.0001

Note: Listing of occupations is in order used in form A; the reverse order was used in form B. Wording of occupational titles is identical to that in questionnaire. Chi-square tests were performed across the entire response distribution (i.e., excellent, good, average, below average, poor, and no opinion); the degrees of freedom for the tests were 5. To conserve space, only the distributions for the excellent response category are shown; this category accounted for a majority of the variability across forms.

NS Not significant.

[a]Computed on assumption that sampling efficiency of clustered example was 66 percent that of equivalent simple random sample. See text for further discussion.

form A, this question was the very first question in the survey. The interviewer was instructed to begin with a standard introduction.

> Hello (respondent's name), I am (interviewer's name) conducting a study for the Caravan Surveys of Opinion Research Corporation of Princeton, New Jersey. In this interview we would like to ask your opinion on a number of different subjects.

The interviewer then proceeded to the special introduction required for the National Science Board's questions,

> I am now going to ask you a group of questions that come from the National Science Foundation, which is a federal agency. They are preparing a report that will discuss public attitudes toward science and technology. Your participation in this survey will be very helpful to them, but it is entirely voluntary. No records will be kept that will allow your individual reply to be associated with you.

The item on the social standing of occupations immediately followed. In form B, the survey began with the same standard introduction, but then proceeded to ask a series of 38 questions on litter, hamburger makers, and hospitalization and medical insurance.[8] Following these questions, the interviewer delivered the NSF introduction and proceeded to ask the question on the prestige of occupations.

The divergence in the results obtained from these two different administrations is striking. Table 6 presents the relevant comparisons. For 8 of 10 occupations, rated prestige is lower when the question is asked at the beginning of the survey (form A). For 7 of the 10 occupations, this difference is 5 or more percentage points, and in four cases, it exceeds 10 percentage points. The sole

---

[8] This survey, the ORC Caravan, consisted of eight parts, each funded by different organizations. The Opinion Research Center treats each section of the questionnaire as the confidential property of the sponsoring organization. Thus, ORC has not been able to make available to us the actual questions used in each section. However, Dean Behrend and his staff have made available a summary of the content of each section and have been helpful in answering questions about survey administration. The following lists derived from summaries prepared by ORC staff describe the content of survey sections.

1. Seriousness of litter problem in U.S.
2. Who is responsible for problem
3. Degree of activity of eight organizations in fighting litter problem
4. Awareness and sponsor identification of antilitter advertising
5. Use of one-way or returnable packaging in purchase of beverages

1. Frequency of serving hamburgers in household
2. How many cooked at one time
3. Ownership and usage of hamburger makers (electrical appliance)
4. Purchase intention and brand intention for a hamburger maker.

1. A series of questions on hospitalization insurance, family members covered, etc.
2. Proportion of expenses paid by such coverages
3. Hospitalization incidence last year
4. Estimated cost of hospitalization

exceptions to this general pattern occur for businessmen and ministers; there the discrepancies are of trivial magnitude (0.0 percent and +0.7 percent).

Clearly, responses to this question were not identical on the two forms of the survey. Respondents who were first exposed to questions on litter, hamburger makers, and hospitalization gave generally more favorable evaluations of these 10 professional occupations. Why this happened is unclear. One might speculate that survey respondents have an initial set against the use of extreme response categories (e.g., excellent). This bias may diminish with practice in responding to survey questions. However, some experimental evidence suggests a modest trend in the opposite direction (cf. Kraut et al., 1975). Alternatively, one might speculate that sequencing banal questions on beverage containers, litter, and hamburger makers, immediately before questions about acute medical problems

Table 7. Effects of Experimental Manipulation of Question Context Upon the Likelihood Respondents Would Express a Great Deal of Confidence

Institution[a]	Proportion expressing a great deal of confidence		Discrepancy	χ^2	p
	Neutral context	Alienation context			
Major companies	.264	.190	−.074	11.1	.0008
Organized religion	.329	.309	−.020	0.6	NS
Education	.294	.284	−.010	0.1	NS
Executive branch of the Federal Government	.126	.133	+.007	0.1	NS
Organized labor	.114	.117	+.003	0.0	NS
Press	.180	.228	+.048	5.0	.025
Medicine	.472	.456	−.016	0.3	NS
TV	.141	.139	−.002	0.0	NS
U.S. Supreme Court	.303	.285	−.015	0.5	NS
Scientific community	.421	.369	−.052	3.8	.05
Congress	.130	.136	+.006	0.1	NS
Military	.314	.299	−.015	0.3	NS
Banks and financial institutions	.351	.317	−.034	1.8	.18

Note: This analysis focuses attention upon the "great deal of confidence" category in accord with common reporting practices (cf. *The Harris Survey*, December 6, 1973; September 30, 1974; October 6, 1975; March 22, 1976; March 14, 1977; January 5, 1978). We have eliminated missing data ("don't know," no answer, etc.) from the response distributions for each item. Chi-square statistics have one degree of freedom and are corrected for continuity (Yates correction). Given that assignment to experimental conditions was fully random, the analysis treats the respondents (average sample size about 1,500) as a universe and tests the hypothesis that the distribution of responses is independent of experimental condition.

Question: I am going to name some institutions in this country. As far as the *people* running them are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them?

NS Not significant.

[a]In order of presentation to respondents.

and experience with doctors and hospitals, created a proscience and proprofessional evaluation bias. The latter speculation may be plausible, especially because respondents were told that they were being asked to evaluate these professional occupations for "the National Science Foundation, which is a federal agency ... preparing a report on public attitudes toward science and technology."

One could, of course, speculate endlessly about the causes of the observed anomaly. It is not our intention to provide a definitive interpretation for the context artifact shown in table 7. That would not be possible given the data at hand. Rather, we wish to know what these results tell us about our initial hypotheses. In this regard, we observe that:

1. The "prestige or general standing" of an occupation is not a well-defined concept.
2. The question requires a somewhat arbitrary choice between response categories (excellent vs. good vs. average vs. below average vs. poor).
3. The concept of the job of businessman, for example, is rather imprecise: Does it mean the local grocer or the president of General Motors?
4. The question is one to which people probably give little thought and it has few behavioral consequences in respondents' everyday lives.

PATTERNS OF ASSOCIATION

Our next set of examples concerns a rather different sort of disagreement. In the preceding sections of this chapter, we have been concerned with whether the univariate patterns of response obtained from different surveys were comparable, for example, did two surveys provide consistent estimates of the level of public support for spending on science and technology. In the present section, we are concerned with whether the *patterns of association* between variables measured in different surveys vary systematically. That is, would one come to the same conclusion about the association between education, for example, and a given attitude—regardless of the survey contexts in which the measurements were made?

In this area, the prevailing wisdom is that even with major wording changes, not to mention context, the bivariate distribution of variables will be undisturbed, even though the marginal (univariate) distributions may vary substantially. As one review of survey research practice recently observed,

> The solution to this problem [of fluctuation in univariate distributions arising from changes in wording] advocated by Davis and other experienced survey investigators is to ignore single variable attitudinal results and concentrate on relationships. The assumption seems to be that single variable distributions vary for reasons that are artifactual, frivolous, or even quite meaningless, but that the ordering of respondents on items—

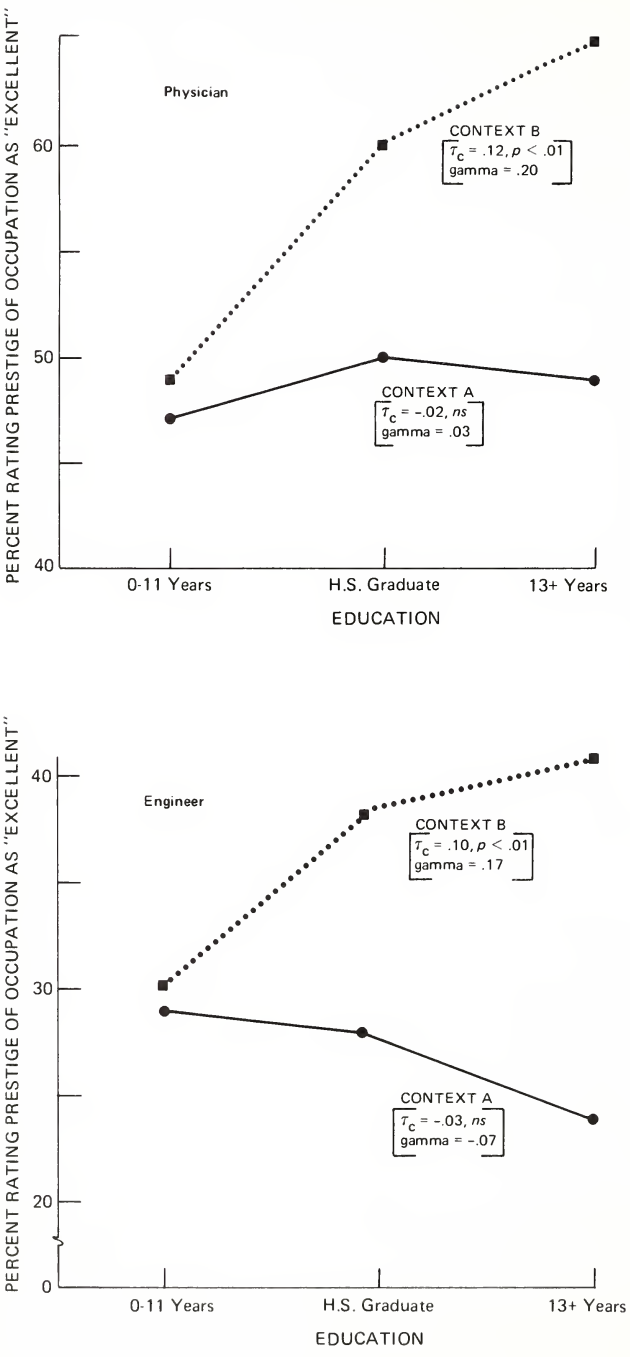
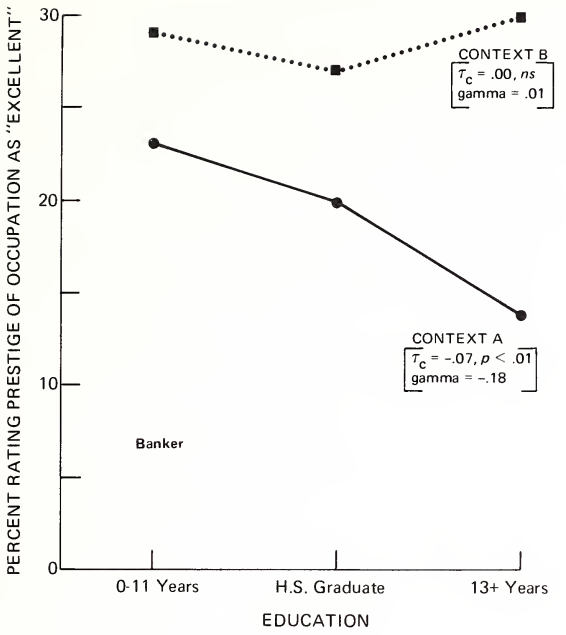


Figure 5. Relationship between ratings of occupational prestige and respondents' educational level for two forms of questionnaire. *Source:* Science Indicators survey, 1976.



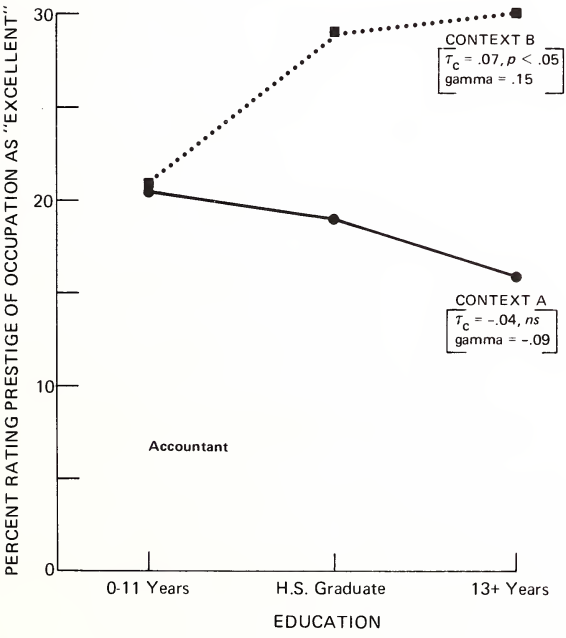


Figure 5. Relationship between ratings of occupational prestige and respondents' educational level for two forms of questionnaire. (Continued)

> and therefore associations among items—are largely immune to this problem. (Schuman and Duncan, 1974, p. 234)

This complacent position has recently been questioned (Schuman and Duncan, 1974; Schuman, 1974; Schuman and Presser, 1977). In their work on the effects of question wording, Duncan and Schuman (1977) concluded that: "[Their research] argues for caution in making the assumption that multivariate patterns of responses will be relatively unaffected . . . even though the univariate response distributions are affected" (p. 9).

Some further information on this question can be gleaned from analysis of the Science Indicators data on occupations. In figure 5 we plot responses to four items from the occupational prestige question by the educational level of the respondents. These results show that the context effects observed in tables 4 and 5 are most pronounced for the highly educated. This, in turn, causes the bivariate patterns of association between respondents' educational levels and their occupational ratings to *vary systematically* between form A and form B. Using ordinal measures of association, we observe modest (median $\gamma = .16$) and generally significant positive correlations between educational level and the likelihood of rating the prestige of these occupations as excellent. In contrast, the four correlations between prestige ratings and education are modestly *negative* in form B of the questionnaire; in one instance (banker), this association is significantly negative ($\gamma = -.18$).

Thus, the conclusions one would reach about the relationship of respondents' education and their evaluation of occupations depends upon the survey context in which the questions were asked. Clearly, for these data the assumption that measurement artifacts are restricted to the univariate response distributions is unwarranted. Even ignoring the rank information, we would still find evidence for an effect upon the multivariate pattern of response. Treating the three categories of education (E) as an unordered classification, the patterns of response (R) to different forms (F) were only poorly fit by models that excluded the three-way interaction term (ERF). In particular, models fitting only the marginals for the (ER) (FR) (EF) distributions produced the following fits to the data:

Occupation	χ^2	p
Engineer	5.4	.07
Physician	4.3	.11
Accountant	5.0	.08
Banker	3.9	.14

(In each case the degrees of freedom associated with the test or fit were 2; computations assume that clustering effects diminished the effective sample size by one-third.)

INCOMPLETE EXPLANATIONS: CONTEXT EFFECTS AND PUBLIC CONFIDENCE

In an earlier paper (Turner and Krauss, 1978), large and persistent discrepancies between Harris and HORC time series on public confidence in the leaders of national institutions were analyzed. In that analysis, a variety of explanations for the discrepancies were investigated and discarded. These explanations included sampling variability, nonrepresentativeness of samples, the untoward effects of quota sampling, and temporal variation in public attitudes. It was concluded that the discrepancies between the Harris and NORC series in their estimates of the level and trends across time in public confidence arose from the effects of large nonsampling errors in those series.

It was speculated that such nonsampling errors might arise, in part, because the questions used to measure public confidence were embedded in rather different survey contexts. These contexts varied both across survey organizations and within organizations across time. Particular attention was drawn to two instances of such contextual variation.

1. In 1976, the Harris questions on public confidence followed a series of negatively worded questions designed to measure political alienation.
2. The order in which particular institutions appeared in the confidence question varied, and in some years particular institutions were presented along with partial repetitions of the question.

It was hypothesized, in particular, that variations in use of the alienation context depressed the levels of confidence found by Harris in 1976 and variations in use of a *people running* prefix accounted for the erratic behavior of NORC's estimates of confidence in organized religion. The latter effect was thought to occur because in some years people responded to the prompt "organized religion," while in other years they were prompted with "how about the people running organized religion."

To provide some experimental evidence upon such context effects, NORC incorporated an experimental manipulation of survey context in the 1978 General Social Survey. Six alienation items were presented either immediately before or immediately after the confidence question. These items were phrased:

> Now I want to read you some things some people have told us they have felt from time to time. Do you tend to feel or note . . .
>
> 1. The people running the country don't really care what happens to you.
> 2. The rich get richer and the poor get poorer.
> 3. What you think doesn't count much anymore.
> 4. You're left out of things going on around you.
> 5. Most people try to take advantage of people like yourself.

6. The people in Washington, D.C. are out of touch with the rest of the country.

It was hypothesized that exposure to these negatively worded alienation items would depress respondents' tendency to report a great deal of confidence in the leaders of the various institutions. The effects of this experimental manipulation are shown in table 7; the order of institutions in this table corresponds to their order of presentation in the survey questionnaire.

What can one conclude from these results? At present, our analyses are in a rather preliminary stage; however, certain things seem clear. First, there is evidence that this variation in context did produce some significant variations in estimates of the proportion of the population having a great deal of confidence. In particular, for the institution that immediately followed the alienation items (major companies) the difference between contexts is −7.4 percentage points. Smaller but still reliable differences were also found for two other institutions (press, +4.8 percent; and scientific community, −5.2 percent). Curiously, while the alienation items generally reduced the frequency of the great-deal-of-confidence response, the reverse effect was found for the press. When measured after a series of items focusing upon political alienation, confidence in the press rose.

A second conclusion we draw from these results is that context of the sort manipulated in this experiment could provide only a partial explanation for the discrepancies observed between the 1976 Harris and NORC estimates. In that year, discrepancies of up to 16 percentage points were observed. In no instance did the experimental manipulation produce discrepancies of this magnitude.

We do note, however, that the NORC experimental manipulation did not fully duplicate the alienation context of the relevant Harris survey. In particular, two questions interspersed between the alienation and confidence questions in the 1976 Harris survey were omitted from NORC's experimental context manipulation. These questions were:

1. Compared to 10 years ago, do you feel the quality of life in America has improved, grown worse, or stayed the same?
2. Compared to 10 years ago, do you feel the leadership inside and outside of government has become better, worse, or stayed the same?

The questions about confidence in the people running national institutions followed these items. The omission of these two questions, particularly the one on leadership, does cause some uncertainty in generalizing from these experimental results to the actual measurements made in 1976.

Patterns of Association

Preliminary examination of these data also revealed evidence that the context manipulation had effects upon the patterns of association between confidence

and other variables. Using the three confidence items that showed significant shifts in their univariate distributions, we examined the relationship of confidence to alienation and to respondents' educational level in order to determine whether there were significant context effects upon the multivariate response distributions. We found little evidence of such an effect for education. Confidence in the press and major companies showed no significant association with education in either form of the questionnaire, while confidence in the scientific community had virtually identical assocation with education level ($\gamma = .30$) in the two questionnaire forms.

Subsequently, we examined the association between the three confidence measures and responses to the three alienation items (1, 2, and 6) that we judged to be most related (semantically) to confidence in national institutions. Figure 6 presents one example of the basic data. Using log-linear techniques (cf. Goodman, 1971, 1972) to model the response distribution of alienation (A) by confidence (C) by questionnaire context (Q), we found some evidence of context effects upon the multivariate distributions. In particular, using a model that was maximally constrained to fit the observed patterns of response but that excluded the three-way interaction term (CAF), we could not obtain an adequate fit to the data in two of nine instances ($p < .05$). And, we obtained a rather poor fit ($p < .20$) in two further instances. Table 8 provides details of these analyses (we report fits only for the maximally constrained noninteractive model (CA, CF, FA) since this is the appropriate comparison model for testing the null hypothesis of no context effect upon the patterns of association).

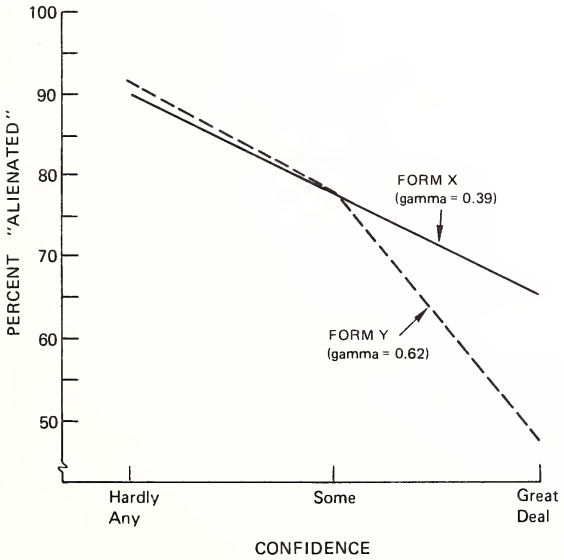


Figure 6. Relationship of confidence in major companies to alienation response for two forms of questionnaire. In form X, the confidence question precedes the alienation items; in form Y, the alienation items precede the confidence question. Alienation item: The rich get richer and the poor get poorer. *Source:* NORC General Social Survey, 1978.

Table 8. Test for Context Effects on Patterns of Association Between Confidence and Alienation

Alienation item	Confidence item	Ordinal association (γ)[a] Form X	Ordinal association (γ)[a] Form Y	Log-linear interaction test[b] χ^2	Log-linear interaction test[b] p
The people running the country don't really care what happens to you.	Major companies	.31	.32	0.0	NS
	Press	.16	.15	0.0	NS
	Scientific community	.18	.30	1.7	.20
The rich get richer and the poor get poorer.	Major companies	.39	.62	7.4	.007
	Press	.09	.13	0.0	NS
	Scientific community	.24	.22	0.1	NS
The people in Washington, D.C. are out of touch with the rest of the country.	Major companies	.21	.39	1.8	.18
	Press	.14	.15	0.9	NS
	Scientific community	.07	.27	4.0	.05

Note: Model fit to response distribution for alienation (A) by confidence (C) by questionnaire context (Q) is a maximally constrained nonsaturated model. In Goodman's notation, it is (CF) (CA) (FA). Failure to fit a model of this type to the data indicates an interaction, that is, that the pattern of association between the variables was not independent of questionnaire context.

NS Not significant.

[a] Form X of the questionnaire presented the confidence questions prior to the alienation items; form Y presented them in reverse sequence.

[b] df = 1.

For discussion purposes, let us briefly consider the instance in which this multivariate context effect is strongest. As figure 6 shows, we found a considerably stronger inverse association between alienation responses (the rich get richer and the poor get poorer) and confidence in the people running major companies when the alienation item follows rather than precedes the confidence question ($\gamma = 0.62$ vs. $\gamma = 0.39$). An examination of the γ coefficients shown in table 8 revealed that this particular relationship holds true in seven of the nine other comparisons. The two exceptions involved reversals of trivial magnitude.

Although our analyses are still at a preliminary stage and the experiment itself imperfectly replicated the actual context variation, we have nonetheless observed significant effects upon both univariate and bivariate response distributions. However, these effects were neither so pervasive nor so overwhelming in magnitude as to provide a *complete* explanation for the discrepancies observed in the 1972-1977 Harris and NORC confidence series. Clearly, many aspects of the behavior of these disagreeable series remain to be understood.

DISCUSSION AND CONCLUSIONS

This chapter was intended to raise many questions but only to hint at answers. Our collection of disagreeable examples was selected to explore some prelimi-

nary hypotheses concerning the types of subjective indicators that are more (and less) vulnerable to artifacts of measurements. Thus, we have tried to assess the degree to which our hypothesized typology was confirmed or disconfirmed by each example.

In addition to providing some preliminary conceptual organization, we hope that the preceding evidence focuses attention upon intersurvey comparability in the measurement of subjective phenomena. The study of disagreeable data is not an end in itself, however. To be useful, it must stimulate the difficult process of ferreting out explanations for particular anomalies and deducing general principles—where they exist. In this regard, we believe that the foregoing evidence allows one to dispose of three common and complacent apologia for inconsistencies in survey estimates of subjective phenomena.

First, because of the range of examples presented here and elsewhere (e.g., Cowan et al., 1978; Gibson et al., 1978; Smith, 1978; Turner and Krauss, 1978), one cannot ascribe these discrepancies to the deficient practices of any particular survey research organization. Such a position would be both unfair and unfaithful to the observed facts. We have observed both discrepancies and consistencies in comparisons involving estimates made by a wide range of research organizations.

Second, both our analyses and those recently undertaken by Duncan and Schuman (1977) and Schuman and Presser (1977) indicate that artifacts in the survey measurement of subjective phenomena are not limited to univariate response distributions. Hence, it is not always safe to assume that analyses focusing on multivariate patterns of association between variables will be resistant to the anomalies encountered in the analysis of univariate distributions.

Last, it appears that no single explanation is likely to be adequate to explain all the observed discrepancies. Experimental data on artifacts in the confidence time series suggest that context, for example, is probably only a partial explanation for the discrepancies in these time series. Various other sources of error (e.g., interviewer effects) need to be considered.

There is much to be learned about the reliable measurement of subjective phenomena. But we need not be discouraged by our confessions of ignorance. New knowledge is sought when accepted truths are found wanting. At present, reality has belied our expectations about the behavior of these data. Our evidence suggests the need for coordinated research designed to improve our understanding of the validity and reliability of survey-based estimates of subjective phenomena. The growing importance of such data in policymaking and the various social indicators enterprises makes such research imperative.

This research is likely to require coordinated studies across research organizations of the error structure of our survey measurements. While some appropriate work can be undertaken by individual investigators or solitary research organizations, there is a clear need to demonstrate the validity and reliability of

comparisons involving data from diverse governmental and nongovernmental sources. Collaborative tests conducted across many survey research organizations would seem an appropriate strategy for calibrating measurement procedures and improving the reliability of estimates of subjective phenomena. In this regard, the procedures developed to assess analytical measurements made by different laboratories in the physical and chemical sciences may offer useful guidance (cf. Youden, 1975; Steiner, 1975; Boffey, 1975).

The cost and difficulty of improving our understanding of the problems affecting survey estimates of subjective phenomena should not be underestimated. However, those who take seriously the tasks of social reporting and the monitoring of changes in the subjective states of the population would seem to have little choice. Such research needs to be done.

Vulnerable Indicators

At the outset, we hypothesized that indicators of some phenomena were more vulnerable to artifacts of measurement than others. In particular, we speculated that measurement artifacts would be more likely to afflict estimates of phenomena that were ambiguous in concept and had little importance for the everyday life of respondents and whose measurement involved a choice between relatively amorphous response categories.

This hypothesis predicts that survey measurements of nonsubjective phenomena, such as chronological age, should be relatively invulnerable to artifacts of measurement (e.g., effects of variant question wordings, survey context). Davis (1976) has studied estimates of the sex, race, age, religion, and educational distribution of the population in 30 sample surveys conducted between 1952 and 1973 by SRC and Gallup. These estimates do, in fact, show good, although not perfect, consistency with one another and with independent estimates made by the Census Bureau and surveys conducted by NORC.

For subjective phenomena, our analysis of eight pairs of estimates of women's fertility expectations revealed a pattern of consistency, which was within the range expected on the basis of sampling fluctuations. Expectations of childbearing, while clearly a subjective phenomenon, are predicted to be less vulnerable to measurement artifacts by our hypothesis. This prediction follows from the fact that the concept of childbearing is unambiguous, the response categories have a clear meaning, and the question itself is directly relevant to the everyday life of the respondents (married women of childbearing age).

The discrepancies observed in our other analyses involved questions whose character is consistent with our preliminary hypotheses. In table 9, we summarize these comparisons together with those presented in three other recent publications (Turner and Krauss, 1978; Smith, 1978; Kalton et al., 1978). We have included only recent studies because the earlier literature was reviewed by Sudman and Bradburn (1974) in their book, *Response Errors in Surveys*. Their

Table 9. Summary of Comparisons of Survey Estimates of Subjective Phenomena

Topic	Response categories	Type of comparison[a]	Comparisons	Significant discrepancies	Maximum discrepancy[b]	
					Raw percentage	*z* score
1. Confidence in people running national institutions	Great deal of confidence, only some confidence, hardly any confidence	Different houses[c]	27	18	16	3.6
		Context experiment	13	3	7	3.3
2. Evaluation of amount of Federal income tax	Too high, about right, too low	Context experiment[c]	1	1	14	5.9
3. General happiness	Very happy, pretty happy, not too happy	Different houses[d]	2	1	15	6.2
		Same house/different surveys[e]	2	1	11	5.5
4. Evaluation of prestige of occupations	Excellent, good, average, below average, poor	Context experiment	10	7	15	5.9
5. Evaluation of contemporary driving standards	Lower than they used to be, higher than they used to be, about the same	Context experiment[f]	2	1	7	3.5
6. Evaluation of traffic noise	Becoming noisier, less noisy, or about the same	Context experiment[f]	2	1	6	2.4
7. Evaluation of need for nighttime truck deliveries to retail stores	Yes, no[g]	Context experiment[f]	2	1	9	≈ 2.8
8. Evaluation of spending on national problems	Too much, about right, too little	Different houses[h]	10	5	13	5.5
9. Misanthropy I: people try to take advantage of you	Try to take advantage if they got a chance, try to be fair	Different houses[h]	2	0	4	1.8
10. Misanthropy II: most of time people try to be helpful	Try to be helpful, mostly just looking out for themselves	Different houses[h]	2	1	6	2.5

(See footnotes on p. 71.)

Table 9. Summary of Comparisons of Survey Estimates of Subjective Phenomena—Continued

Topic	Response categories	Type of comparison[a]	Comparisons	Significant discrepancies	Maximum discrepancy[b]	
					Raw percentage	*z* score
11. Misanthropy III: most people can be trusted	Can be trusted, can't be too careful in dealing with people	Different houses[h]	2	1	7	3.3
12. Courts' treatment of criminals	Deal too harshly or not harshly enough with criminals	Different houses[h]	1	0	1	0.6
13. Men or women better suited for politics	Men better suited, equally suited, women better suited	Different houses[h]	1	0	6	1.8
14. Vote for woman for President	Yes, no	Different houses[h]	1	1	5	2.2
15. Children safe bicycling in local area	Yes, no	Context experiment[f]	2	0	2	≈ 0.6
16. Law requiring police permit to purchase gun	Favor, oppose	Different houses[h]	3	0	2	0.6
17. Support for legalization of marijuana	Yes, no	Different houses[h]	1	0	3	1.5
18. Political party with which one identifies self	Republican, Democrat, independent, other	Different houses[h]	3	1	4	3.2

19. Favor death penalty for murder	Yes, no	Different houses[h]	3	0	3	1.5
20. Ideal number of children for family	0, 1, 2, 3, . . .	Different houses[h]	1	0	1	0.6
21. Expected number of children	0, 1, 2, 3, . . .	Different houses[h]	8	1	9	2.0

Note: Comparisons are restricted to those cited in four recent reviews (Kalton et al., 1978; Smith, 1978; Turner and Krauss, 1978; this chapter). Only comparisons involving measures of subjective phenomena are included; behavioral measures (e.g., voting) and demographic estimates are excluded. Comparisons involving variant question wording have been excluded—except where noted otherwise. All comparisons made within 5 months of each other.

[a]Comparisons involve either (1) independent estimates made by different survey research organizations (different houses); (2) experiments involving the manipulation of the context in which survey questions were presented (context experiment); or (3) estimates derived from separate surveys conducted by the same survey research organization (same house/different surveys).

[b]Discrepancies are presented as absolute difference between percent of population giving selected response (raw percentage P), and the z score for the observed discrepancy,

$$z = \frac{P_1 - P_2}{\hat{\sigma}_{P_1 - P_2}}$$

where $\sigma_{P_1 - P_2}$ is standard error for expected value of difference between two indpendent estimates, discrepancies with z scores greater than 2.0 can be considered reliable (i.e., $p < .05$).

In computing these statistics, we have used standard error estimates from the source publications, if available. When data were derived from clustered samples and no estimates of sampling error were available in the source publication, we computed estimates based on the formula for a simple random sample, but we employed deflated sample sizes ($N' = 0.66N$) to allow for design inefficiencies.

[c]Turner and Krauss (1978).

[d]Comparison involves a modest difference in question wording, see page 44.

[e]Comparisons were made by pooling estimates from 2 months of NORC Continuous National Survey closest to dates of General Social Surveys.

[f]Kalton et al. (1978).

[g]Source does not quote explicit response categories; these are implied by text of question.

[h]Smith (1978).

conclusions, however, are somewhat at variance with those that have emerged from later work.

The summary presented in table 9 shows a rough correspondence with the typology of our preliminary hypotheses. Thus, we find relatively more significant discrepancies for items 1-14. These items involve measurements of rather amorphous concepts such as confidence in the people running institutions and evaluations of occupational prestige, national spending, and contemporary driving standards. These same items also require a choice between relatively imprecise response categories, such as:

1. Great deal, only some, hardly any (confidence)
2. Too much, about right, too little (spending)
3. Excellent, above average, average, below average, poor (prestige)

In contrast, questions 15-21 about the legalization of marijuana, gun control, the death penalty, political party affiliation, and fertility yield relatively few discrepancies. The latter questions involve somewhat less amorphous topics and response categories, for example, support or nonsupport of legislation, the name of an actual political party, or an actual number of children expected.

One further difference between these two groups of questions is the (likely) salience of the topics for respondents. The latter group of questions inquires about topics that have been the subject of considerable public discussion (e.g., capital punishment and the legislation of marijuana) or that are connected with specific behaviors that have tangible behavioral components (e.g., voting and party registration or contraceptive practice). For this group of questions, only two significant discrepancies were observed in 19 comparisons. In contrast, the first group of items asked about topics that, we suspect, would not be subject to considerable public discussion as phrased in these questions, that is, Do I have confidence in the people running major companies? What is the prestige of accountants? Are we spending too much on science? Over one-half of the comparisons (42 of 82) involving this group of questions produced significant discrepancies.

Although we believe that a case can and has been made for the typology we hypothesized, table 9 is not without its counterexamples. Voting for a woman president (yes or no) is a concrete action and related issues have been the subject of considerable public discussion. Yet, in the one instance where a comparison was possible, we find a modest (4 percent) but reliable difference between estimates derived from independent surveys using this question. Thus, while the evidence is generally compatible with our hypotheses, the correspondence between our typology and the available data is less than perfect.

Other evidence supports our typology of vulnerable indicators. In an unpublished experiment, Duncan and Schuman (1977) found context-induced variations in responses to five (of seven) questions. These questions were generally consistent with our typology; they measured, for example, agreement with

statements such as

> Public officials really care about what people like me think.
>
> Given enough time and money, almost all of man's important problems can be solved by science.

Respondents chose among four response categories: strongly agree, agree, disagree, strongly disagree.

In addition, recent studies (Cowan et al., 1978; Gibson et al., 1978) of survey-based estimates of crime victimization have revealed that substantial and significant variations in respondents' reports of crime were induced by differences in questionnaire contexts. While the National Crime Surveys (NCS) incorporate rather specific descriptions in their questions, for example,

> Did anyone beat you up, attack you, or hit you with something, such as a rock or bottle (In the last 6 months)?

crime victimization rates for each of the 13 cities surveyed by NCS show a strong rise in crime reporting when measurements of victimization are made after a series of attitudinal questions on crime and fear of crime. Gibson et al. (1978) report that this measurement artifact produced a relative increase of 12 percent in the rate of property crime and 21 percent in the rate of personal crime.

Examination of the NCS questionnaire suggests that this result fits within the rubric of our typology. The victimization questionnaire includes, for example, the following items:

> 42. Did anyone try to attack you in some other way (other than any incidents already mentioned)?
> 46. Did you find any evidence that someone attempted to steal something that belonged to you (other than any incidents already mentioned)?

The concepts of *attempted* theft and attempted attack, in these questions, are somewhat amorphous. Our hypotheses would predict that estimates derived from these questions would be more vulnerable to artifact than those derived from questions on the actual incidence of theft and physical assault. We also suspect that these items bear a major responsibility for the observed variations in crime victimization rates (remember, *attempted* theft or assault is also a crime). However, even questions on actual assault are not without their ambiguities, for example, when does a friendly poke become a hostile blow?

Analysis of the NCS data by Cowan et al. (1978) provides some confirmation of our predictions. For example, simple assault and aggravated assault show little evidence of context-induced artifacts (z scores for difference in rates: 0.3 and

0.2). However, attempted assault with a weapon and attempted simple assault show much stronger evidence of context-induced artifacts ($z = 1.8$ and $z = 2.8$). Similarly, robbery estimates show less evidence of artifacts than attempted robbery ($z = 0.2$ vs. $z = 1.2$), and estimates of personal larcencies involving actual contact between victim and thief are less affected by context than estimates of larcenies involving no personal contact ($z = 0.3$ vs. $z = 3.4$).

FUNDAMENTALS AND FUTURE DIRECTIONS

We think it would be a mistake to view the problems presented by these disagreeable examples as strictly methodological. Similarly, we do not think that remedies should be sought in narrowly conceived methodological research.

We suggest that there is a need for a reconsideration of the psychological assumptions that underlie the practice of survey research. We do not have a particular agenda of research to propose in this area, rather we believe that a fundamental reconsideration of the psychological foundations of survey research ought to be encouraged, and we applaud those independent initiatives that have recently emerged (e.g., Nisbett and Wilson, 1977; Wilson and Nisbett, 1978; Fischoff et al., 1979).

It should be obvious that there is a fundamental relationship between psychological concerns and the practice of survey research. While this is doubtlessly a truism, it is often ignored.

The most fundamental phenomena of survey research are quintessentially psychological in character. They arise from a complex interpersonal exchange, they embody (or are contaminated by, if you wish) the subjectivities of both interviewer and interviewee, and they present their interpreter with an analytical challenge that requires a multitude of assumptions concerning, among other things, how respondents experience the reality of the interview situation, decode the meaning of survey questions, and respond to the social presence of the interviewer and the demand characteristics of the interview.

In this regard, we note that the average user of social survey data knows little or nothing about the interviewers who are the other half of the social interaction that produces these data. While few survey research organizations would fail to provide routine demographic information on respondents, similar information is seldom—if ever—provided about interviewers. Thus by default, interviewers are treated by most analysts as anonymous and passive encoders of the subjective reality of respondents. It is a bit odd that as social scientists we must adopt such a narrow view of the social realities involved in our own work.

The burden of the observed anomalies should prompt a reconsideration of the psychological foundations of survey research. The foregoing examples are indicative of the deficient state of our present knowledge, and we hasten to note that such topics have not been the subject of particularly active research in the last decade. We doubt that there are any instant solutions. However, it also seems clear that complacency will not suffice.

REFERENCES

Abrams, M. 1973. "Subjective Social Indicators." In United Kingdom, Central Statistical Office, *Social Trends: 1973.* London: HMSO.

Andrews, F., and S. Withey. 1976. *Social Indicators of Well-Being: Americans' Perceptions of Life Quality.* New York: Plenum Press.

Blalock, H. 1972. *Social Statistics,* 2d ed. New York: McGraw-Hill, 1972.

Blau, P., and O. D. Duncan. 1967. *The American Occupational Structure.* New York: Wiley.

Boffey, P. 1975. "Scientific Data: 50 Percent Unusable; Widespread Defects in Laboratory Work Found by National Bureau of Standards." *Chronicle of Higher Education,* February 24, 1975, 1.

Bradburn, N. 1969. *The Structure of Psychological Well Being.* Chicago: Aldine.

Campbell, A., P. Converse, and W. Rodgers. 1976. *The Quality of American Life: Perceptions, Evaluations and Satisfaction.* New York: Russell Sage.

Caplan, N., and E. Barton. 1976. *Social Indicators 1973: A Study of the Relationship of the Power of Information and Utilization by Federal Executives.* Ann Arbor: Institute for Social Research.

Cowan, C., L. Murphy, and J. Weiner. 1978. "Effects of Supplemental Questions on Victimization Rates from the National Crime Surveys." Paper presented at the 138th Annual Meeting of the American Statistical Association, San Diego, August 14-17.

Davis, J. 1975a. "Communism, Conformity, Cohorts and Categories: American Tolerance in 1954 and 1972-3." *American Journal of Sociology,* 81: 491-513.

____. 1975b. "Does Economic Growth Improve the Human Lot? Yes Indeed About .0005 Per Year." Paper presented to the International Conference on Subjective Indicators of the Quality of Life, Cambridge, England.

____. 1976. "Background Characteristics in the U.S. Adult Population 1952-1973: A Survey-Metric Model." *Social Science Research* 5: 349-383.

Duncan, O. D. 1961. "A Socioeconomic Index for All Occupations." In A. Reiss, ed., *Occupation and Social Status.* New York: Free Press.

____. 1972. "Federal Statistics, Non-Federal Statisticians." *Proceedings of the American Statistical Association (Social Statistics Section).* 152.

____. 1979. "Indicators of Sex Typing." *American Journal of Sociology* 85: 251-260.

Duncan, O. D., and H. Schuman. 1977. *An Experiment on Order and Wording of Questions.* Unpublished manuscript, Department of Sociology, University of Arizona.

Easterlin, R. 1974. "Does Economic Growth Improve the Human Lot? Some Empirical Evidence." In P. Davis and M. Reder, eds., *Nations and Households in Economic Growth.* New York: Academic Press.

Fischoff, B., P. Slovic, and S. Lichtenstein. 1979. "Knowing What You Want to Know: Measuring Labile Values." In T. Wallsten, ed., *Cognitive Processes in Choice and Decision Behavior.* Hillsdale, N.J.: Earlbaum.

Freedman, R., A. Hermalin, and M. Chang. 1975. "Do Statements About

Expected Family Size Predict Fertility? The Case of Taiwan, 1967-1970." *Demography* 12: 407-416.

Gibson, C., G. Shapiro, L. Murphy, and G. Stanko. 1978. "Interaction of Survey Questions as It Relates to Interviewer-Respondent Bias." Paper presented at the 138th Annual Meeting of the American Statistical Association, San Diego, August 14-17.

Goldberg, D., H. Sharp, and R. Freedman. 1959. "The Stability and Reliability of Expected Family Size Data." *Millbank Memorial Fund Quarterly* 37: 369-385.

Goldfield, E., A. Turner, C. Cowan, and J. Scott. 1977. "Privacy and Confidentiality as Factors in Survey Response." *Proceedings of the American Statistical Association (Social Statistics Section).* 1-11.

Goodman, L. 1971. "The Analysis of Multidimensional Contingency Tables." *Techonometrics* 13: 33-61.

____. 1972. "A General Model for the Analysis of Surveys." *American Journal of Sociology* 77: 1035-1086.

Gurin, G., J. Veroff, and S. Feld. 1960. *Americans View Their Mental Health.* New York: Basic.

Hall, J., and D. Jones. 1950. "The Social Grading of Occupations." *British Journal of Sociology* 1: 31-55.

Ho, C. Y., R. W. Powell, and P. E. Liley. 1974. "Thermal Conductivity of the Elements: A Comprehensive Review." *Journal of Physical and Chemical Reference Data* 3 (suppl. 1): 1-244.

Hunter, J. S. 1977. *Quality Assessment of Measurement Methods.* In National Academy of Sciences, *Environmental Monitoring*, vol. 4a, Washington, D.C.: National Academy of Sciences, National Research Council.

Kalton, G., M. Collins, and L. Brook. 1978. "Experiments in Wording Opinion Questions." *Applied Statistics* 27: 149-161.

Kish, L. 1965. *Survey Sampling,* 2d ed. New York: Wiley.

Kraut, A., I. Wolfson, and A. Rothenberg. 1975. "Some Effects of Position on Opinion Survey Items." *Journal of Applied Psychology* 60: 774-776.

Martin, E. 1978. "Trends in Victimization: Problems of Measurement." Paper presented at the 86th Annual Meeting, American Psychological Association, Toronto, August 28-September 1.

____. 1981. "A Critique of Replication Studies of Social Change: Problems in Monitoring Change." In P. Rossi and J. Wright, eds., *Handbook of Survey Research.* New York: Academic Press (in press).

Mason, K., J. Czaijka, and S. Arber. 1976. "Changes in U.S. Women's Sex Role Attitudes." *American Journal of Sociology* 41: 573-598.

National Academy of Sciences, Committee on National Statistics-National Research Council. 1979. *Privacy and Confidentiality as Factors in Survey Response.* Washington, D.C.: National Academy of Sciences.

National Science Board. 1973. *Science Indicators: 1972.* Washington, D.C.: U.S. Government Printing Office.

____. 1975. *Science Indicators: 1974.* Washington, D.C.: U.S. Government Printing Office.

____. 1977. *Science Indicators: 1976.* Washington, D.C.: U.S. Government Printing Office.

Nisbett, R., and T. Wilson. 1977. "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84: 231-259.

Rich, R. 1975. *An Investigation of Information Gathering in Seven Federal Bureaucracies: A Case Study of the Continuous National Survey.* Unpublished doctoral dissertation, University of Chicago.

Schuman, H. 1974. *Old Wine in New Bottles: Some Sources of Response Error in the Use of Attitude Surveys to Study Social Change.* Paper prepared for Research Seminar in Quantitative Social Science, University of Surrey (England), April 1974.

Schuman, H., and O. D. Duncan. 1974. "Questions About Attitude Survey Questions." *Sociological Methodology: 1973-4.* San Francisco, Jossey-Bass.

Schuman, H., and M. Johnson. 1976. "Attitudes and Behavior." In A. Inkeles, ed., *Annual Review of Sociology* 2: 161-207.

Schuman, H., and S. Presser. 1977. "Question Wording as an Independent Variable in Survey Analysis." *Sociological Methods and Research* 6: 151-170.

Sewell, W., and R. Hauser. 1975. *Education, Occupation, and Earnings.* New York: Academic Press.

Sheldon, E. 1971. "Social Reporting for the 1970's." In Presidential Commission on Federal Statistics, *Federal Statistics,* vol. 2. Washington, D.C.: U.S. Government Printing Office.

Smith, T., 1978. "In Search of House Effects: A Comparison of Responses to Various Questions by Different Survey Organizations." *Public Opinion Quarterly* 42: 443-463.

____. 1979. "Happiness: Time Trends, Seasonal Variations, Intersurvey Differences, and Other Mysteries. *Social Psychology Quarterly* 42: 18-30.

Staines, G., and R. Quinn. 1979. "American Workers Evaluate the Quality of Their Jobs." *Monthly Labor Review* 102: 3-12.

Steiner, E. 1975. *Planning and Analysis of the Results of Collaborative Tests.* Washington, D.C.: Association of Official Analytical Chemists.

Sudman, S., and N. Bradburn. 1974. *Response Effects in Surveys.* Chicago: Aldine.

Treiman, D. 1977. *Occupational Prestige in Comparative Perspective.* New York: Academic Press.

Turner, C., and E. Krauss. 1978. "Fallible Indicators of the Subjective State of the Nation." *American Psychologist* 33: 456-470.

United Nations, Department of Economic and Social Affairs. 1975. *Toward a System of Social and Demographic Statistics.* New York: United Nations.

U.S. Department of Commerce. 1973. *Social Indicators, 1973.* Washington, D.C.: U.S. Government Printing Office.

____. 1977. *Social Indicators, 1976.* Washington, D.C.: U.S. Government Printing Office.

Waksberg, J. 1975. "How Good Are Survey Statistics?" *Proceedings of the American Statistical Association (Social Statistics Section),* 26-27.

Wilson, F., and L. Bumpass. 1973. "The Prediction of Fertility Among Catholics." *Demography* 10: 591-597.

Wilson, T., and R. Nisbett. 1978. "The Accuracy of Verbal Reports About the Effects of Stimuli on Evaluations and Behavior." *Social Psychology* 41: 118-131.

Youden, W. 1975. *Statistical Techniques for Collaborative Tests.* Washington, D.C.: Association of Official Analytical Chemists.

Youden, W., and E. Steiner, eds. 1975. *Statistical Manual of the Association of Official Analytical Chemists.* Washington, D.C.: Association of Official Analytical Chemists.

Irregularities in Survey Data

Angus Campbell
Institute for Social Research
University of Michigan

it is always difficult to evaluate the kind of analysis Dr. Turner presents. Do we take his examples of irregularity as isolated incidences having only case history significance, or do they deserve a broader interpretation? Turner finds these discrepancies to have weighty implications, and his chapter gives the impression that the reality of survey data is considerably seamier than our innocent expectations.

Turner describes his analysis as being at "a preliminary stage," and one cannot properly complain that he has not done more than he intended. I am concerned, however, that an unsophisticated reader will take his "disagreeable examples" as a general indictment of the quality of survey data, and I regret that the language of the article rather invites the reader to that conclusion. I feel confident that by a judicious selection of examples I could present a set of data that would make survey findings appear to be as solid as Gibraltar, but it would have the same basic weakness as Turner's piece. Generalities cannot be derived from isolated and purposefully selected incidents.

Turning to the author's first example of where surveys disagree, we find a very curious selection of data. The argument revolves around figure 1, Trends in Self-reported Happiness, 1971-1973. There appears in the *Social Psychology Quarterly* in March 1979, 2 years before the present publication, an article by Tom W. Smith entitled "Happiness: Time Trends, Seasonal Variations, Inter-survey Differences, and Other Mysteries." It deals with precisely the same data that Turner reviews but extends them in time from 1971 to 1977. Smith presents a more extensive review of the context problem and also shows a rather persistent seasonal variation. When these effects are removed, it appears that the NORC and SRC data show a generally similar pattern over the period 1971-1977, although the NORC percentages of people reporting themselves very happy are consistently a little higher than those of the SRC. The reader who sees only figure 1 in Turner's article would surely conclude that the data of the two organizations were heading off in opposite directions with no relationship

to each other. I hope that anyone interested in the reliability of happiness scores will look up Tom Smith's article.

I would also point out to anyone with such interests that the happiness question is one of the least reliable measures we have for the assessment of psychological well-being. It has all the weaknesses of a single-item measure. It persists in current research because it is the only measure of its kind that has any time depth. George Gallup began asking a question about personal happiness in 1946, and the precise wording that both NORC and SRC have been using originated in 1957. We have so few trend data in this field that we cannot afford to throw anything away, even if it is less reliable than the multi-item measures we now depend on.

One cannot argue with Turner's implication that NORC was ill-advised to place a happiness-in-marriage question directly prior to its question on general life happiness. He demonstrates very effectively, as Smith does in greater detail, that it produced a context effect. I am not prepared, however, to accept his suggestion that happiness is such a "notably amorphous" concept that it cannot be measured reliably. It must be remembered that the experience of happiness is subject to variation resulting from changing external circumstances. Even if it could be measured without error, there would be variation from one measure to the next. In 1978 the SRC reinterviewed 694 people we had interviewed in a national sample in 1971. The correlation of the answers to the standard happiness question over the 7-year period was .36, but the correlation between the answers to an eight-item measure of general affect used in the two surveys was .51. Considering the number of life-shaking events that must have happened to these people over these years, I find the degree of constancy shown in the latter measure to be rather impressive.

Turner's second example subjects NORC to another test from which it emerges quite handsomely. The striking fact about figure 3 is the stunning stability in year-to-year Census reports of number of women expecting no more children. Would that we always had samples of 4,000 to follow change over time. I am surprised that Turner did not raise the question of why the two sides of figure 3 are so similar. Considering the fact that the respondent was asked first, "Do you expect to have any more children?" It is obvious that the proportion who are not ever going to have more children must be the same proportion who are not going to have any in the next 5 years. Indeed, although table 3 does not show contingency instructions, one would assume that questions 2 and 3 were not asked of women who answered question 1 by saying they did not expect any more children. Figure 3*b* is in fact superfluous.

The section "Science and the Public" deals with two rather dramatic examples of context effects in surveys done for the National Science Board. They are impressive and they clearly resulted in the waste of public money. Turner approaches this argument by setting up what appears to me to be a preposterous straw man to the effect that "the prevailing wisdom among survey research-

ers" tells us that question position has no effect on question responses. To support this proposition he presents the quotation from Sudman and Bradburn (1974) on response effects. This quote is certainly very delicately selected; in the next paragraph in their book the authors write, "The findings reported here confirm the analysis of a limited number of studies focusing on question order effects (Bradburn and Mason). That analysis failed to show any consistent order effects, although individual studies did report significant order effects.... Considerably more research will have to be done before we can formulate any theory on position effects."

I don't pretend to represent the prevailing wisdom, but I would think that most survey practitioners with any sophistication about interview construction have learned to be wary about question order. They are not always able to predict the effect of question placement and are sometimes surprised that it is larger or smaller than they expected. But certainly if one is setting up a series of studies to measure trends where the integrity of the absolute values obtained is critical, one makes every effort to keep the question order precisely the same from one wave to the next. Any departure invites the kind of discrepancy that appears in this section of Turner's chapter. I am not sure what theory would have given us any basis to predict that intervening questions on hamburgers and litter would influence estimates of occupational prestige, but it would not have taken a lot of imagination to guess that compelling respondents to consider how effective they expected science and technology to be in solving current problems might influence their views of how much money ought to go into these efforts. The only thing that surprises me about this presentation is how poorly the National Science Board apparently is served by the people who advise it on survey research.

Turner returns again at the end of this section to his argument that context influence is greatest in questions dealing with poorly defined concepts. This hypothesis is not unreasonable to be sure, but one observes that the discrepancies in table 4 in regard to such apparently precise concepts as reducing crime, treating drug addiction, improving the safety of automobiles, and the like are generally larger than those in table 6 involving occupational prestige. The critical factor appears to be the content of the first question and how it relates to the content of the next question.

Turner suggests that the response categories (excellent, good, average, etc.) used in the occupational ratings are arbitrary and therefore subject to measurement error. Psychological scales clearly do not have the precision of measures based on cardinal numbers, but with all their faults they often behave with remarkable discipline. Table 5 presents the marginals from a series of studies of public evaluations of the money spent on various public problems. The percentages march across the page in an impressive display of stability. The only one that does not records the growing support for military expenditures, which Congress now finds irresistible. The labels "too little, about right, and too much" do not appear less arbitrary or ill-defined than excellent, good, average, and the like.

In "Patterns of Association," Turner presents an interesting example of contextual influence not only on marginal distributions but on correlations of these distributions with an outside variable. College-educated respondents were clearly more attentive to the questions they were asked than the less educated respondents, many of whom, we may assume, were expressing what Philip Converse calls nonattitudes. One would have to guess that there is a great deal of random variation in these latter responses, which is generally offsetting, and that the standard deviation of these responses is larger than that of the college graduates even though their mean score is more stable.

It is rather wry that figure 5 shows the patterns it does while Turner is arguing that questions "to which people probably give little thought" are the most susceptible to context influence. It seems quite clear that in this case the respondents who could be expected to be giving the least thought to the question were practically impervious to the influence of the question order. I do not think Turner's argument is totally without merit, but if respondents are presented a series of questions they do not understand, it does not seem likely that one would have much influence on their reactions to the next.

In his 1978 article with Krauss, Turner undertook to explain the rather substantial discrepancies existing in a series of NORC and Harris surveys asking respondents to express their degree of confidence in the leaders of certain national institutions. The data were far from ideal for his purpose; two of the five pairs of surveys were taken at an interval of several months, and there was some variation in the order of questions in the NORC surveys. No simple explanation of the discrepancies in the two sets of data came out of this analysis, and Turner proposed that context effects might account for some part of them. He was particularly taken with the possibility that a set of questions intended to measure political alienation, which preceded the confidence-in-leadership questions in the 1976 Harris survey, had deflated the levels of confidence expressed in that survey.

Perhaps in response to Turner's article, NORC undertook the experiment in 1978 described in "Incomplete Explanations" in his chapter. The impressive fact about the results of this study, as shown in table 7, is the total absence of any sensible pattern in the results that might be attributed to the presence of the alienation questions. Of the 13 responses, 1 seems to have changed clearly beyond the range of chance and in the expected direction. Two others changed enough to suggest an influence, but one changed positively and the other negatively. Despite Turner's belief that this was an instance "where one might anticipate substantial context effects to occur," they did not appear. Perhaps if the two additional Harris questions that NORC omitted had been included, his anticipations might have been more satisfactorily fulfilled, perhaps not. My guess is that if one searched about, one could probably find a number of examples of this kind—the questionnaire appears to be loaded for context effect, but the actual influence is either invisible or illogical. As Sudman and Bradburn concluded some 16 years ago, order effects tend to be rather inconsistent.

Turner tells us in his conclusions that his intention is "to raise many questions but only to hint at answers." If I felt confident that his readers would take his examples merely as hints, I would feel more comfortable with his presentation, but in fact I am not sure that he takes them as such himself. He appears to feel that he has "disposed" of house effects as a contributor to discrepancies in survey data, although his evidence would seem quite inadequate for this achievement. He leaves the impression that context effects are a major contributor to survey variability, although after devoting most of his presentation to them he admits they may be only a partial explanation. He suggests fallibility in measures of subjective phenomena that surely requires a much broader documentation than he has given it. It is probably virtually inevitable that a collection of case studies will give an impression of general truth even though, as in this case, the author insists that he hopes merely to provide "initial hypotheses around which research may be organized."

At various points, Turner speaks of a need for "coordinated research" to improve understanding of the validity and reliability of survey data. This need has been present ever since World War II, and it becomes increasingly pressing as time passes. I regret that he did not offer any specific suggestions on what a program of coordinated research would look like.

I would like very much to see a serious inquiry into the question of the reliability of survey measures of both objective and subjective phenomena. This inquiry would require a broad-scale review of data from identical questions asked of comparable samples over some period of time. Its purpose would not be to find intriguing examples of consistent or inconsistent data but to plot the variability of the total corpus of data, to compare its distribution to what we would expect from sampling error alone, and to identify those influences that contribute to unreliability.

My own inclination would be to begin this research by dividing these data by agency of origin. Turner makes it clear, both in this chapter and in his 1978 paper (Turner and Krauss, 1978), that he does not regard house differences as very significant. I believe the differences in the quality of the data produced by the Census Bureau, for example, and by the poorest of the commercial polls are substantial, and I regret the tendency of the general public and many scholars to assume that all sample surveys are equally reliable. These agencies differ, not only in their basic sampling designs, but in the precision with which they carry out these designs, in their attention to supervision, and in their general insistence on high standards at all stages of the survey process.

We need a broad-gauge study for which the surviving organizations, private and public, make available all of their trend data over the last 20 years, that is, all distributions resulting from identical questions asked of comparable samples over that period of years or some part of it. The analysis of these data would answer several questions. First, it would tell us what the range of variability in survey data actually is rather than what we would predict it to be on the basis of the formula for sampling error. Second, it would reveal the differences in

reliability that may or may not exist in different kinds of data, objective and subjective, for example. And third, it would make it possible for the first time to know whether the money, time, and care that some survey organizations put into full-scale probability sampling produce more reliable data than those produced by less expensive methods. This would be a very worthwhile undertaking, and the Survey Research Center would be pleased to be the first to make its files available.

No one who is experienced in the generation and analysis of survey data can be unaware of the irregularities that occur in these data. We have all been confronted with outliers that defy explanation. Most of us would agree that even after 35 years of postwar data gathering, survey researchers still have a good deal to be modest about. But survey research must be judged not only by its frailties but by its strengths as well, and the record of achievement is obviously substantial. Research on survey methodology is moving forward at various points around the country, and we may hope that in due course these inquiries will provide answers to some of the questions that now confound us.

REFERENCES

Smith, T. 1979. "Happiness: Time Trends, Seasonal Variations, Intersurvey Differences and Other Mysteries." *Social Psychology Quarterly* 42: 18-30.

Sudman, S., and N. Bradburn. 1974. *Response Effects in Surveys: A Review and Synthesis.* Chicago: Aldine.

Turner, C., and E. Krauss. 1978. "Fallible Indicators of the Subjective State of the Nation." *American Psychologist* 33: 456-470.

Patterns of Disagreement: A Reply to Angus Campbell

Charles F. Turner
National Research Council
National Academy of Sciences

It is, of course, a privilege to have the benefit of Dr. Campbell's comments. There is, it seems, fundamental agreement between us concerning the modest and unsystematic nature of our present understanding of the nonsampling components of variability in survey measurements of subjective phenomena. Although we seem to be in agreement on this basic point, there are some disagreements between us.

ON EVIDENCE

Dr. Campbell wishes to dismiss the examples presented in "Surveys of Subjective Phenomena" and elsewhere (Turner and Krauss, 1978) as a mere set of case studies. He implies that by the judicious selection of contrary examples, he might demonstrate that such survey measurements are "as solid as Gibraltar."

Assessing the representativeness of any set of examples is an important scientific labor. In the present instance, it is an admittedly difficult task. It is, nonetheless, unfortunate that Dr. Campbell has chosen to rest his argument upon speculation rather than enriching our discussion by presenting evidence in support of his claim.

"Surveys of Subjective Phenomena" does present, in total, 101 recent examples of replicated measurements (table 9), and it reviews their conformance to a general typology suggested earlier by Turner and Krauss (1978, note 12). Doubtlessly, this compilation does not include every recent instance of replicated measurements. It does, nonetheless, represent a substantial number of them. In fact, it is relatively uncommon for two survey organizations to ask precisely the same question at the same point in time. Indeed, one often finds that allegedly identical questions turn out to be worded differently when one

I would like to thank Elizabeth Martin and Theresa DeMaio for their helpful comments on an earlier draft of this manuscript.

consults the actual survey questionnaires. For example, NORC and SRC's happiness measurements (analyzed by Campbell et al., 1976, p. 26; Andrews and Withey, 1976, p. 319; Rodgers and Converse, 1975, p. 130) actually involved questions that differ slightly in wording (although these authors treat them as equivalent without noting this difference).

Because wording differences sometimes have substantial effects on responses, we included only instances in which identically worded questions were asked within 3 months of each other.[1] With one exception,[2] our compilation did not knowingly exclude any substantial body of recent replications. Being a first step, however, we did not perform an exhaustive search but attempted to incorporate only recent published work readily accessible to us. The 101 replicated measurements presented in table 9 of "Surveys of Subjective Phenomena" include all instances reported in the recent work of Smith (1978), Turner and Krauss (1978), and Kalton et al. (1978). Furthermore, drawing as it does upon Smith's (1978) compilation, our analysis includes every known contemporaneous replication of any of the subjective items asked in the six NORC General Social Surveys conducted between 1972 and 1977. In addition, some replicated measurements that were not incorporated among these 101 examples were discussed elsewhere in the text, for example, Census Bureau research on the National Crime Survey measurements (cf. Cowan et al., 1978).

As for the judiciousness of our selection of examples and Dr. Campbell's fear that we invite readers to conclude that a "general indictment of the quality of survey data" is warranted, I would merely ask readers to recall the second example we chose for extended discussion. This example involved measurements of women's fertility expectations. As Dr. Campbell notes, these measurements are quite well behaved. They were selected for early and extended discussion in order to preclude the misreading of our work as a general indictment of all survey measures of subjective phenomena. Our intent is made explicit in our caveat:

> Lest the reader be misled by our first example, we hasten to note that we do not believe that all survey measurements of subjective phenomena are equally vulnerable to artifactual biases. Rather, we wish to delineate areas in which artifact-induced discrepancies might be expected and the factors likely to cause such misbehavior. (p. 48)

---

[1] In this regard, we also note that while Turner and Krauss (1978), report the results of 45 replications involving five surveys conducted by NORC and five by Louis Harris and Associates, our summary excluded two-fifths of those replications. As Dr. Campbell correctly observes, there was a considerable time interval between the NORC and Harris measurements in some years. In the original review (cf. Turner and Krauss, 1978, pp. 458-459 and table 4), separate analyses were done for replications involving overlapping surveys versus surveys that did not overlap in time. It was found that "considering only the [overlapping] measurements, we found that, on the average, discrepancies during this period are marginally greater than those in [nonoverlapping] years" (p. 459). Nonetheless, the nonoverlapping measurements were not included in table 9.

[2] The only substantial body of replicated measurements that we knowingly excluded was the presidential popularity measurements. Those data were not readily available to us at the time; a comprehensive review of those measurements is presently underway.

In conclusion, I suggest that a dispassionate reading of the text does not support Dr. Campbell's attempt to dismiss the evidence as a judiciously selected set of case studies.

ON SURVEY PRACTICES

Dr. Campbell also makes a considerable point of objecting to the claim that question context is often (inappropriately) presumed to have no effect on measurements. He asserts that "certainly if one is setting up a series of studies to measure trends where the integrity of the absolute values obtained is critical, one makes every effort to keep the question order precisely the same from one wave to the next. Any departure invites the kind of discrepancy that appears in this section [of 'Surveys of Subjective Phenomena']." Perhaps, most survey researchers would acknowledge Dr. Campbell's advice as correct, but in fact, they frequently disregard it. Omnibus surveys designed to track trends frequently change their content so that, in fact, questionnaire content and question order are variable over time. Neither the NORC General Social Survey, the SRC omnibus surveys, nor most commercial surveys (e.g., Harris and Gallup) replicate questionnaires in their entirety. Thus, changes over time are confounded with changes that may occur because of variations in questionnaire content and question order. The fact that survey researchers frequently fail to replicate question context from one survey to the next supports the inference that they do not regard the possible effects as very important.

Explicit statements of this presumption and some empirical evidence in support of it may be found in the work of commercial pollsters (e.g., Clancy and Wachsler, 1971).[3] Moreover, prevailing practice, even in scientific papers, does not ordinarily require the allowance that intersurvey comparisons confound population change with changes in survey questionnaires (or other survey procedures). Thus, one seldom encounters analyses expressing the concerns of Duncan and Evers (1975) that

> The study design does not permit us to measure the influence of the prior question sequence on responses to the woman's work questions, [and thus] we cannot rule out an intersurvey difference in frame of reference. (p. 133)

Indeed, two of the leading researchers on subjective measures of well-being (Andrews and Withey, 1976, p. 319, footnote 6), have themselves commented on the same NORC (GSS) happiness data presented in figure 1 and table 2 of "Surveys of Subjective Phenomena." These authors observed that the NORC data show "a sharp and unexpected rise" in 1973-1975. These authors did not recognize or allow for the potential context artifacts that Dr. Campbell believes to be apparent in the NORC (GSS) series.

---

[3] Anecdotal evidence suggests, however, that a few survey organizations have adopted priority systems for ordering questionnaire topics in order to provide some standardization of question contexts across surveys.

In this regard, I would also refer readers to the distressing experience of the Census Bureau's National Crime Survey (discussed briefly in "Surveys of Subjective Phenomena" and more extensively in Cowan et al., 1978, and Gibson et al., 1978). Despite careful consideration and planning, even the most experienced survey research organizations have sometimes found themselves surprised by large and unexpected variations in survey measurements that were induced by context variations initially thought to be unproblematic.

In summary, I suggest that the real problems in this area are not caused by carelessness or a lack of sophistication but rather our lack of theoretical knowledge about the nonsampling sources of variance in our measurements.

SOME QUIBBLES

Having discussed Dr. Campbell's major reservations, I should like to touch briefly on some minor points:

1. *On fertility estimates.* Dr. Campbell correctly points out that a woman's fertility expectations in *all future years* are not independent of her expectations for the *next 5 years*. At a minimum, knowing that a woman expects no further children, one also knows that she expects to have no children in the next 5 years (barring coding errors or logical inconsistency). However, Dr. Campbell errs when he says that the two measurements are redundant. While there is some overlap, unique information is contained in each measurement. For example, knowing that a woman expects no children in the next 5 years does not rule out the possibility that she expects to have children at a later age. Because each measurement contains some unique information, they are not entirely redundant. For this reason, we chose to present both series.

2. *On house effects.* Variations in surveys measurements arise from variations in survey procedures not from the mere fact that a survey is done by organization X rather than organization Y.[4] The aim of scientific research should be to identify those survey procedures that induce variability in measurements.

   Discussions of house effects (i.e., the residual variability in measurements associated with the organization doing the measurement) are not helpful in attempts to improve survey measurements. Such discussions do not identify the *sources* of measurement variability that affect comparisons of data produced by different houses, and thus, they offer no guidance on how to provide better (across house) standardization of measurements.

---

[4] Except for the special case where the mere auspices of the survey induce variations in response. Slight evidence of such effects was found in surveys conducted for the Committee on National Statistics (National Academy of Sciences, 1979) by the Census Bureau and the Survey Research Center (Michigan). Such effects were found for questions that asked (directly or indirectly) about the organizations themselves (e.g., the trustworthiness of data produced by different types of survey organizations). This special case might be defined as a pure house effect.

3. *On sampling.* Dr. Campbell observes that there is a need for assessing the impact of various sampling strategies on survey measurements. In "Surveys of Subjective Phenomena," we purposely focused attention on measurement variability induced by nonsampling factors. This does not imply that we believe sampling factors induce no variability in measurements or that the failure to draw proper samples is not, itself, a problem. Sampling, however, is an area in which there is a well-developed theory and an understanding among users of these data as to what constitutes a good sample.

   Many surveys draw inadequate samples (see, Bailar and Lanphier, 1978); however, the dangers of such deficiencies are well known, and if care is exercised, such deficiencies need not occur.[5] When deficiencies do occur, it is not for want of an adequate theory of sampling. This is not the case when one is dealing with the variability induced by nonsampling factors. In this regard, our present situation is remarkably similar to that described by Dr. Campbell 35 years ago:

   > At the present time the most highly developed aspect of surveying procedures appears to be the sampling. This is not to imply that all surveys, or even most surveys, employ well designed and unbiased sampling methods. It is true, however, that the science of sampling has reached the point that it can select unbiased samples of known probable error to represent virtually any universe a surveyer is apt to be interested in. In contrast, the interviewing phase of survey technology is by no means as well controlled and it is likely that in many surveys the interviewing error is considerably larger than the sampling error. (Campbell, 1946, p. 65)

4. *On contexts and confidence.* In his comments on the confidence experiment, Dr. Campbell correctly observes that the pattern of results in this 1978 General Social Survey experiment was not what we had predicted. What had been predicted was that all confidence measurements would be depressed when measured after the series of alienation items. What we found was that responses to one question were substantially affected (7.4-percentage-point change) and that two other questions evidenced more modest effects.

   While Dr. Campbell observes no pattern in our results, I would point out that the question showing a clear effect in the predicted direction was the one *immediately* following the alienation items. Contiguity, one suspects, plays an important part in determining when one question affects the responses to another.

---

[5] Furthermore, for those, like Dr. Campbell, who are concerned about the untoward impact of compromised sampling, (e.g., the use of quotas rather than full-probability methods at the household level), there are easily applied checks. Indeed, the checks that have been done suggest that for many questions the use of quotas at the household level does not necessarily yield notably different estimates than full-probability methods (Turner and Krauss, 1978, p. 462; Stephenson, 1979).

5. *On history.* Dr. Campbell implies that my own work on the happiness measurements merely rediscovers things already reported by Smith (1979). In fact, my findings, first presented at the 1978 convention of the American Psychological Association, relate to work begun in 1977[6] and originally discussed at an informal meeting attended by several social scientists in the fall of that year (footnote 5, p. 44).

   Tom Smith and I have worked both independently and collaboratively on this and related topics[7] for the past 2 years. I, too, would recommend that readers consult his writings on this subject. Interested readers will find that Smith's suspicions in regard to the 1972-1974 happiness measurements are the same as mine and that his conclusions regarding seasonal variations in these measurements are considerably more cautious than Dr. Campbell's assertion that these measurements show a "rather persistent seasonal variation." Smith writes, "In sum, the hypothesis that happiness (and conceivably other measures of global well-being) follows a seasonal rhythm is plausible, but not proven" (1979, p. 27).

6. *On the need for broad-gauged studies.* Dr. Campbell urges upon us the need for broad-gauged studies of the reliability of survey measurements. On this point, we are in complete agreement.

   Some important work in this direction was initiated by the American Statistical Association (cf. Bailar and Lanphier, 1978), and other important studies have been recently undertaken by individual investigators (e.g., Schuman and Presser, 1981; Bradburn and Sudman, 1979). In the same vein, the National Science Foundation has recently funded a major review of the uses, reliability, and meaningfulness of survey measurements of subjective phenomena. This work is being undertaken by a panel of statisticians and social scientists chaired by O. D. Duncan, under the auspices of the Committee on National Statistics of the National Research Council (National Academy of Sciences). Hopefully, this and similar ventures will go some way toward meeting the important and mutually agreed upon goal of improving both survey measurements and our understanding of their properties and problems.

IN CONCLUSION: ORDERLINESS AMIDST ERROR

In "Surveys of Subjective Phenomena," we argued: (1) that nonsampling factors, such as question context, can induce substantial variations in survey

---

[6] The history of the present publication apparently is typical of most scientific writing. Garvey's (1979) review suggests that 90 percent of scientific findings appear in informal communications prior to publication. Formal publications, in turn, often lag 4-5 years behind informal communications in most disciplines.

[7] Other research on related topics (e.g., Schuman and Presser, 1981; Duncan and Schuman, 1980) also suggests that nonsampling artifacts may be more problematic than previously assumed and that understanding these artifacts is particularly crucial for those wishing to construct time series of subjective social indicators.

measurements of subjective phenomena, and (2) that certain types of survey questions are particularly vulnerable to such measurement artifacts. For example, based on the evidence we reviewed, it appeared that general or summary questions were more vulnerable to such artifacts than specific questions. If the context in which a question is embedded provides a framework that respondents use in decoding the question's meaning, then this result is intuitively reasonable. General (or vague) questions are more labile in their meaning, and thus, they are more in need of interpretation. Hence, question context may play a relatively greater role in determining responses to such questions.

As part of an ongoing program of secondary analysis and new experimentation, we have subsequently carried out three pilot experiments using the general happiness question and the more specific marital happiness item discussed in "Surveys of Subjective Phenomena." Measurements were made using identially worded questions[8] in surveys by the *Washington Post* (May 1979), the Survey Research Center of Michigan (August 1979) and the National Opinion Research Center (February-March 1980). In each survey the order of question presentation was experimentally manipulated. In (approximately) one-half of the interviews,[9] the general happiness item followed the question on marital happiness (*controlled context*); in the remaining half of the interviews, it followed whatever else was in that section of the survey questionnaire (*uncontrolled context*). Since the three surveys were otherwise completely different, the question immediately preceding the general happiness item varied *across surveys* in the uncontrolled context. So, for example, in the *Washington Post* survey, the general happiness item followed a question on income in the uncontrolled context, while in the SRC survey it followed a series of items on the gas shortage and a question asking marital status and in the NORC survey a series of five questions asking "how much satisfaction" the respondent gets from city or place of residence, hobbies, family life, friendships, health and physical condition.

A parallel variation in measurement context occurred for the marital happiness item. In (approximately) one-half of the interviews it followed the question on general happiness (controlled context), and in the remainder of the interviews, it *preceded* the general happiness item and thus followed anything else in that section of the survey questionnaire (same items noted above).

Figure 1 displays the proportion of respondents saying they were very happy in response to the general and marital happiness question in each experiment.[10]

---

[8] General happiness: Taken all together, how would you say things are these days—would you say that you are very happy, pretty happy, or not too happy? Marital happiness: Taking things all together, how would you describe your marriage? Would you say that your marriage is very happy, pretty happy, or not too happy?

[9] In the SRC and *Post* experiments, the sample division was approximately 50:50; in the NORC experiment, it was approximately 66:33.

[10] All samples include only married respondents since the marital happiness question could not be asked of unmarried respondents. Samples are drawn from the adult (18 years and

(Continued)

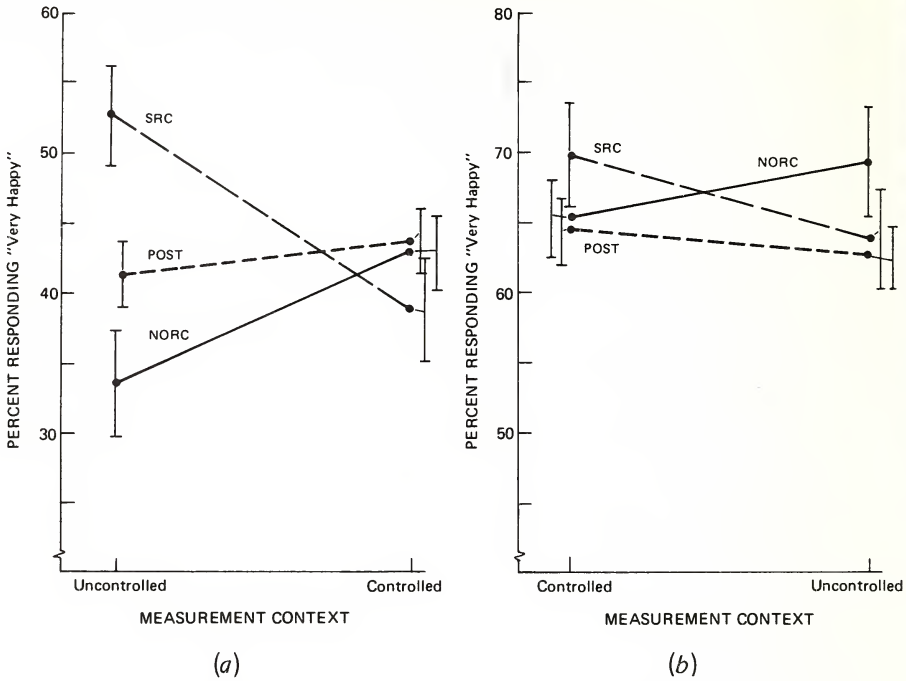


Figure 1. Responses to questions on (*a*) general happiness and (*b*) marital happiness. Error bars demark ±1 standard error for estimates. Data are from surveys conducted by the Survey Research Center, University of Michigan (SRC), the National Opinion Research Center (NORC), and the George Fine Organization for the *Washington Post*. All samples were restricted to married respondents who had telephones in their residences.

While, the results of these experiments are perplexing in some regards, the data do consistently support the hypothesis suggested in "Surveys of Subjective Phenomena." Thus, we note that the more specific question about marital happiness yields equivalent estimates in all three surveys and in both measurement contexts (χ^2 = 10.8; df = 10; p = .37).[11] Secondly, we observe that the

(Continued)

older) population of the continental United States. The SRC and *Post* experiments were done using telephone interviews and random digit-dialing samples; the NORC experiment was done in a face-to-face survey using a multistage area probability sample (see NORC, 1980, for description of sample design). Assignment of respondents to experimental conditions was done at random. Selection of individual respondents in households in the NORC and SRC surveys was done by random selection from all eligible members of the household; in the *Post* survey, selection was from those eligible members of the household who were at home at the time of the initial contact.

All data are unweighted. Due to an error in the preliminary processing of the SRC data, six eligible respondents were eliminated from the tabulation. This error changes none of the SRC results by more than 1 percent.

Data from the NORC experiment include only respondents who reported that they had a telephone in their residence in order to make the sample comparable to those of SRC and the *Washington Post*. (Respondents (circa 3 percent of sample) who refused to give their telephone numbers to the interviewer and thus were not asked the location of their phones were included in this tabulation.)

[11] Likelihood ratio chi-square fit to model 1 (table 1).

general happiness question does yield consistent estimates *but only when the measurement context is controlled.* When the measurement context is left uncontrolled, the measurements vary from 33.5 to 52.7 percent very happy.

Fitting alternative log-linear models (table 1) to the cross-tabulation of happiness response (H) by survey (S) by measurement context (C), we find that a three-way interaction term (HSC) is required to fit the data for the general happiness estimates. In contrast, for the marital happiness measurements, a model (table 1, model 1: H, S, C) positing that the measurements are independent of both context and the survey organization doing the measurements provides an adequate fit to the data (p = .37).

While these results fit the pattern hypothesized in "Surveys of Subjective Phenomena," it is important to note that the direction and magnitude of the artifacts found in the general happiness measurements would have been difficult to predict *in advance*. Thus, it is hard to intuit why SRC's measurements in the uncontrolled context are so much higher than those of NORC and the *Post*. Nonetheless, since the different organizations' measurements agree in the controlled context, we can effectively rule out organizational differences in sampling, processing, and so on, as an explanation for the disagreements we have observed.

These results again suggest that our disagreeable survey measurements are not random noise. Rather, disagreements in our survey measurements have identifi-

Table 1. Test of Alternative Models for Behavior of Happiness Measurements

Model	Marginals fit	df	χ^2	p
General happiness measurements				
1. Stable measurements	(H) (CS)	10	28.0	.002
2. Context effect	(HC) (CS)	8	27.7	.001
3. Survey effect	(HS) (CS)	6	20.8	.002
4. Context and survey effects	(HS) (HC) (CS)	4	20.0	.001
5. Interaction effect	(HSC)	0	0.0	(X)
Marital happiness measurements				
1. Stable measurements	(H) (CS)	10	10.8	.37
2. Context effect	(HC) (CS)	8	10.7	.21
3. Survey effect	(HS) (CS)	6	5.9	.44
4. Context and survey effects	(HS) (HC) (CS)	4	5.5	.24
5. Interaction effect	(HSC)	0	0.0	(X)

Note: Models were fit using procedures developed by Goodman (1971). χ^2 values are likelihood ratio chi-square statistics. Variables included in this analysis are:

H = response to happiness question (three categories: very happy; pretty happy; not too happy). Respondents who did not answer this question (1 percent or less) were excluded from sample.
S = survey (three categories: NORC; SRC; *Washington Post*).
C = measurement context (two categories: controlled context; uncontrolled context).

X Not applicable.

able causes, and it appears that there are systematic differences between the types of questions that are more (or less) vulnerable to measurement artifacts. If this is true, some improvements in survey practices may follow. At a minimum, in the present case we know that special caution needs to be exercised with general questions on happiness. Replications that do not control the context of such measurements appear to be especially vulnerable to contamination by substantial nonsampling artifacts.

In conclusion, I would suggest that the search for general principles that will allow us to identify and control for the vulnerabilities of our survey measurements are an important part of the work of the next generation of survey researchers. The growing desire to track long term changes in the subjective state of the Nation (proposed elsewhere by Campbell et al., 1976, and Duncan, 1969) requires an improved understanding of the error structure of measurements. This need grows daily with the expansion of the corpus of subjective measurements. The ability to confidently disentangle true change in the population from that induced by changes in our measuring instruments is a prerequisite of reliable inference. On this point, Dr. Campbell and I appear to speak with one voice, and I echo his hope that future research on these questions "will provide answers to some of the questions that now confound us."

REFERENCES

Andrews, F., and S. Withey. 1976. *Social Indicators of Well-Being: Americans' Perceptions of Life Quality.* New York: Plenum Press.

Bailar, B., and C. Lanphier. 1978. *Development of Survey Methods to Assess Survey Practices: A Report of the American Statistical Association's Pilot Project on the Assessment of Survey Practices and Data Quality in Surveys of Human Populations.* Washington, D.C.: American Statistical Association.

Bradburn, N., and S. Sudman. 1979. *Improving Interview Method and Questionnaire Design.* San Francisco: Jossey-Bass.

Campbell, A. 1946. "A Summing-Up." *Journal of Social Issues* 2: 58-67.

Campbell, A., P. Converse, and W. Rodgers. 1976. *The Quality of American Life: Perceptions, Evaluations and Satisfactions.* New York: Russell Sage.

Clancy, K., and R. Wachsler. 1971. "Positional Effects in Shared Cost Surveys." *Public Opinion Quarterly* 35: 258-265.

Cowan, C., L. Murphy, and J. Weiner. 1978. "Effects of Supplemental Questions on Victimization Rates from the National Crime Surveys." Paper presented at the 138th Annual Meeting of the American Statistical Association, San Diego, August 14-17.

Duncan, B., and M. Evers. 1975. "Measuring Change in Attitudes Toward Women's Work." In K. Land and S. Spilerman, eds., *Social Indicator Models.* New York: Russell Sage.

Duncan, O. D. 1969. *Toward Social Reporting: Next Steps.* New York: Russell Sage.

Duncan, O. D., and H. Schuman. 1980. "Effects of Question Wording and

Context: An Experiment with Religious Indicators." *Journal of the American Statistical Association* 75: 269-275.

Garvey, W. D. 1979. *Communication: The Essence of Science.* New York: Pergamon Press.

Gibson, C., G. Shapiro, L. Murphy, and G. Stanko. 1978. "Interaction of Survey Questions as It Relates to Interviewer-Respondent Bias." Paper presented at the 138 Annual Meeting of the American Statistical Association, San Diego, August 14-17.

Goodman, L. 1971. "The Analysis of Multidimensional Contingency Tables." *Technometrics* 13: 33-61.

Kalton, G., M. Collins, and L. Brook. 1978. "Experiments in Wording Opinion Questions." *Applied Statistics* 27: 149-161.

National Academy of Sciences, Committee on National Statistics-National Research Council. 1979. *Privacy and Confidentiality as Factors in Survey Response.* Washington, D.C.: National Academy of Sciences.

Rodgers, W., and P. Converse. 1975. "Measures of the Perceived Overall Quality of Life." *Social Indicators Research* 2: 127-152.

Schuman, H., and S. Presser. 1981. *Questions and Answers: Experiments in the Form, Wording, and Context of Survey Questions.* New York: Academic Press.

Smith, T. 1978. "In Search of House Effects: A Comparison of Responses to Various Questions by Different Survey Organizations." *Public Opinion Quarterly* 42: 443-463.

Smith, T. 1979. "Happiness: Time Trends, Seasonal Variations, Intersurvey Differences and Other Mysteries." *Social Psychology Quarterly* 42: 18-30.

Stephenson, C. B. 1979. "Probability Sampling with Quotas: An Experiment." *Public Opinion Quarterly* 43: 477-496.

Turner, C., and E. Krauss. 1978. Fallible Indicators of the Subjective State of the Nation." *American Psychologist* 33: 456-470.

Subjective Indicators of Neighborhood Quality

Donald C. Dahmann
Bureau of the Census

INTRODUCTION

Current assessments of the quality of life in the United States focus not only on the levels of achievement for basic concerns such as food and shelter but also on aspects of life that are less tangible and therefore more difficult to measure such as satisfaction and fulfillment from various life experiences and one's sense of self-actualization. This expanded conceptualization of the quality of life underpins developments that have occurred during the last decade in the area of social indicators and social accounting (compare U.S. President's Research Committee on Social Trends, 1933, with Andrews and Withey, 1976; Bell, 1969; Olson, 1969; Sheldon and Moore, 1968; and U.S. Department of Health, Education, and Welfare, 1970). As one group of researchers has stated, "It is no longer enough for the nation to aspire to material wealth; the experience of life must be stimulating, rewarding, and secure" (Campbell et al., 1976).

Quality of life is viewed here as comprised of the collective assessment made by individuals of the various domains of the life experience. Social relations, including family life, marital relations, friendships, organizational affiliations, and the like, represent one bundle of these life experience domains. A second cluster accounts for participation in the economy, including overall standard of living and work both within and outside the home. Personal health represents an independent domain. Another cluster encompasses various aspects of the residential environment, which consists of a series of domains at scales ranging from the dwelling unit through neighborhood (including social and physical entities), community of residence (rural, town, city, suburb, etc.), subnational region (the Rockies, New England, etc.), and the Nation itself (Campbell et al., 1976).[1]

[1] The notion that residence (or housing) is a multidimensional concept was developed by Isler (1970) and has been carried into data presentations and analyses (Dahmann, 1979; Frieden and Solomon, 1977; and U.S. Bureau of the Census, 1978).

Concern for the quality of life at the neighborhood level is longstanding and has focused (generally independently) on either the social or the physical environment. The social aspects of life at the neighborhood or local community scale represent one of sociology's dominant themes (Park, 1916; Gans, 1962, 1967; Suttles, 1968; Hunter, 1974; Stein, 1960) and was the major focus of the recent effort of the U.S. National Commission on Neighborhoods (1979). Research on the physical environment at the neighborhood level, while also dating from the turn of this century (Russell Sage Foundation, 1909-1914), has developed more slowly and increased dramatically only with the general increase in concern for the physical environment during the past decade. The past 10 years have witnessed the development of both objective indicators of local conditions (Antunes and Plumlee, 1977; Berry et al., 1974, 1976; Hoch, 1972; Schmid, 1974; U.S. Council on Environmental Quality, 1970) and subjective indicators measuring reactions to local conditions (Goldblatt, 1977; Lansing and Marans, 1969; Marans and Rodgers, 1976; Michelson, 1966; Newman and Duncan, 1979; Zehner, 1971).

This study focuses on data that currently serve as indicators of the quality of life at the neighborhood level in annual reports to the Congress. In this role, these indicators serve as the principal basis for congressional oversight of progress toward the environmental portion of the national housing goal: a decent home in a *suitable environment* for all American families (U.S. Department of Housing and Urban Development, 1979b). Careful consideration must therefore be paid this particular set of indicators as they form the major statistical benchmark for congressional action on residential environment, that is, the neighborhood.

The indicators of neighborhood quality used in these annual reports derive from data of the Annual Housing Survey (AHS). Data from the 1976 Annual Housing Survey are used to establish which local conditions are of greatest concern to Americans and by inference to identify conditions producing the lowest quality residential environments. The following section introduces pertinent information about the Annual Housing Survey. Responses to the various aspects of the neighborhood surveyed in 1976 are formed into single response indices that rank neighborhood conditions in three types of residential locations—cities and suburbs of metropolitan areas and nonmetropolitan areas (defined as of 1970 in U.S. Office of Management and Budget, 1976). The set of responses used to form this index is disaggregated to form multiple-response measures to identify differences between the perception and evaluation dimensions of subjective responses to neighborhood conditions. The chapter concludes with an evaluation of the information provided by the several approaches to creating subjective indicators of conditions at the neighborhood level.

DATA SOURCE: THE ANNUAL HOUSING SURVEY

The Annual Housing Survey is a large-scale general purpose survey of the Nation's housing that is conducted through household interviews by the Bureau

of the Census for the Department of Housing and Urban Development. Authorized by the Housing and Urban Development Act of 1970 and initiated in 1973, this survey provides information on a wide variety of characteristics of the Nation's existing housing stock, housing costs, and households. The survey consists of a national sample of 70,000-80,000 housing units and samples of 5,000 or 15,000 housing units in each of 60 individual standard metropolitan statistical areas (U.S. Department of Housing and Urban Development, 1979a).

The data examined here were obtained in household interviews conducted between October and December of 1976 in the survey's national sample. This sample includes households from the Census Bureau's 461 primary sampling units, comprised of 923 counties and independent cities located throughout the 50 States and the District of Columbia. The approximately 75,500 households eligible for interviews in 1976 were derived of households in the 1975 sampling frame plus additions to account for the construction of new dwelling units during the intervening year. The original sample (taken in 1973) was drawn from units enumerated in the 1970 Census of Population and Housing (adjusted for new construction) and represented an overall sampling ratio of about 1 in every 1,366 of the Nation's housing units.

The figures in this chapter are derived of estimates of the Nation's total housing inventory as of 1976, which included about 80.9 million housing units in total, 79.3 million year-round units, and 74 million occupied units (which also serves as the total number of households). The three-stage ratio estimation procedure used to determine sample weights and the reliability of the survey estimates are described in the Census Bureau's *Current Housing Reports* for 1976 (U.S. Bureau of the Census, 1978).

ASSESSMENTS OF NEIGHBORHOOD QUALITY

The Annual Housing Survey contains a series of questions concerning specific neighborhood characteristics and an assessment of overall neighborhood quality. In 1976, these questions asked residents to provide (1) an assessment of overall satisfaction with their current neighborhood; (2) evaluations of the quality of a half dozen local public services such as convenience shopping, public transportation, and schools; (3) perceptions of the presence of 12 specific conditions in their neighborhoods; and (4) evaluations of the nuisance level of conditions they reported as existing in their neighborhoods (figure 1).

The 12 conditions surveyed in 1976 may conveniently be divided into four groups: (1) the condition of local structures, (2) the status of local streets or highways, (3) public safety, and (4) pollution. Information on qualities of the local built environment was provided by questions about abandoned or boarded-up structures; rundown occupied dwellings; and industrial, commercial, or other nonresidential activities. The status of local streets was surveyed with questions on state of repair, traffic volume, and the adequacy of lighting. In probing respondents' sense of public security, they were asked to give reactions to crime "on their streets." The several forms of pollution to which reactions

were obtained include noise pollution from both street traffic and aircraft, air pollution as indicated by the presence of odors, smoke, or gas, and the existence of refuse in the form of trash, litter, and junk, either on streets or in open neighborhood lots.

Assessments of these 12 conditions were obtained by asking a set of three questions in hierarchical order (figure 1). Respondents were first asked if each of the 12 conditions existed in their neighborhoods, and for each condition that was reported as existing, respondents were asked if they found the condition to be bothersome. If a condition was assessed as bothersome, respondents were asked if it was "so bothersome that the household would like to move from the neighborhood."

The first of these three questions provides a measure of the cognitive perception of the presence of conditions, that is, whether respondents recalled that the surveyed conditions were present in their local environment. The survey provides no objective measures of the existence and intensity of conditions,

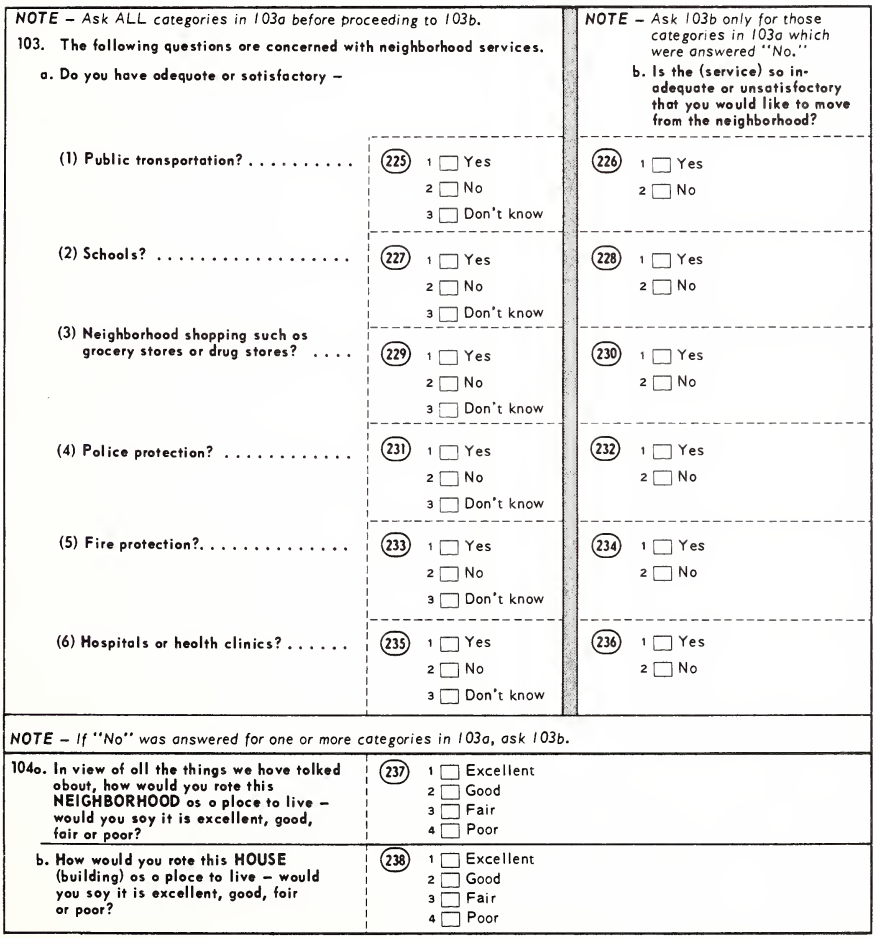
NOTE – *Ask all categories in 102a before proceeding to 102b.*		***NOTE*** – *Ask 102b only for those categories in 102a which were answered "Yes."*	
102a. The following questions are concerned with different aspects of your **PRESENT** neighborhood. Here is a list of conditions which many people have on their streets. Which, if any, do you have?		b. Does the (condition) bother you?	c. Is it so objectionable that you would like to move from the neighborhood?
(1) Street or highway noise?	(213) 1 ☐ Yes 2 ☐ No *	(1) 3 ☐ Yes – Ask c 4 ☐ No	5 ☐ Yes 6 ☐ No
(2) Heavy traffic?	(214) 1 ☐ Yes 2 ☐ No *	(2) 3 ☐ Yes – Ask c 4 ☐ No	5 ☐ Yes 6 ☐ No
(3) Streets or roads continually in need of repair, or open ditches?	(215) 1 ☐ Yes 2 ☐ No *	(3) 3 ☐ Yes – Ask c 4 ☐ No	5 ☐ Yes 6 ☐ No
(4) Roads impassable due to snow, water, etc.?	(216) 1 ☐ Yes 2 ☐ No *	(4) 3 ☐ Yes – Ask c 4 ☐ No	5 ☐ Yes 6 ☐ No
(5) Poor street lighting?	(217) 1 ☐ Yes 2 ☐ No *	(5) 3 ☐ Yes – Ask c 4 ☐ No	5 ☐ Yes 6 ☐ No
(6) Neighborhood crime?	(218) 1 ☐ Yes 2 ☐ No *	(6) 3 ☐ Yes – Ask c 4 ☐ No	5 ☐ Yes 6 ☐ No
(7) Trash, litter, or junk in the streets (roads), or on empty lots, or on properties in this neighborhood?	(219) 1 ☐ Yes 2 ☐ No *	(7) 3 ☐ Yes – Ask c 4 ☐ No	5 ☐ Yes 6 ☐ No
(8) Boarded-up or abandoned structures?	(220) 1 ☐ Yes 2 ☐ No *	(8) 3 ☐ Yes – Ask c 4 ☐ No	5 ☐ Yes 6 ☐ No
(9) Occupied housing in rundown condition?	(221) 1 ☐ Yes 2 ☐ No *	(9) 3 ☐ Yes – Ask c 4 ☐ No	5 ☐ Yes 6 ☐ No
(10) Industries, businesses, stores, or other nonresidential activities?	(222) 1 ☐ Yes 2 ☐ No *	(10) 3 ☐ Yes – Ask c 4 ☐ No	5 ☐ Yes 6 ☐ No
(11) Odors, smoke, or gas?	(223) 1 ☐ Yes 2 ☐ No *	(11) 3 ☐ Yes – Ask c 4 ☐ No	5 ☐ Yes 6 ☐ No
(12) Noise from airplane traffic?	(224) 1 ☐ Yes 2 ☐ No *	(12) 3 ☐ Yes – Ask c 4 ☐ No	5 ☐ Yes 6 ☐ No
NOTE – *If "Yes" was answered for one or more of the categories in 102a, ask 102b.*			

Figure 1. Questions on neighborhood quality from the 1976 Annual Housing Survey.

such as measures of traffic volume in vehicle per hour, roughness of roads, or decibel readings of automobile- or aircraft-generated noise so that it is impossible to determine the actual quantity of conditions that existed in the neighborhoods of the surveyed households.

The second and third questions provide evaluative information. The first asked if the household was bothered by each condition that was reported to exist. The second evaluation question incorporates the concept of mobility as a measure of extreme dissatisfaction and thus serves to anchor subjective reactions to conditions (Hirschman, 1970; Newman and Duncan, 1979).

The national estimates of the number of households reporting that conditions existed, were bothersome, and were extremely bothersome are shown in table 1. The conditions most widely reported were all street related, namely, noisy



NOTE – Ask ALL categories in 103a before proceeding to 103b.

103. The following questions are concerned with neighborhood services.

a. Do you have adequate or satisfactory –

NOTE – Ask 103b only for those categories in 103a which were answered "No."

b. Is the (service) so inadequate or unsatisfactory that you would like to move from the neighborhood?

	103a	103b
(1) Public transportation?	(225) 1 ☐ Yes 2 ☐ No 3 ☐ Don't know	(226) 1 ☐ Yes 2 ☐ No
(2) Schools?	(227) 1 ☐ Yes 2 ☐ No 3 ☐ Don't know	(228) 1 ☐ Yes 2 ☐ No
(3) Neighborhood shopping such as grocery stores or drug stores?	(229) 1 ☐ Yes 2 ☐ No 3 ☐ Don't know	(230) 1 ☐ Yes 2 ☐ No
(4) Police protection?	(231) 1 ☐ Yes 2 ☐ No 3 ☐ Don't know	(232) 1 ☐ Yes 2 ☐ No
(5) Fire protection?	(233) 1 ☐ Yes 2 ☐ No 3 ☐ Don't know	(234) 1 ☐ Yes 2 ☐ No
(6) Hospitals or health clinics?	(235) 1 ☐ Yes 2 ☐ No 3 ☐ Don't know	(236) 1 ☐ Yes 2 ☐ No

NOTE – If "No" was answered for one or more categories in 103a, ask 103b.

104a. In view of all the things we have talked about, how would you rate this **NEIGHBORHOOD** as a place to live – would you say it is excellent, good, fair or poor?	(237) 1 ☐ Excellent 2 ☐ Good 3 ☐ Fair 4 ☐ Poor
b. How would you rate this **HOUSE** (building) as a place to live – would you say it is excellent, good, fair or poor?	(238) 1 ☐ Excellent 2 ☐ Good 3 ☐ Fair 4 ☐ Poor

Figure 1. Questions on neighborhood quality from the 1976 Annual Housing Survey. (Continued)

Table 1. Household Reactions to Neighborhood Conditions: 1976

Condition	Presence		Bothersome		Extremely bothersome	
	Number (thous.)	Percent of all households[a]	Number (thous.)	Percent of all households[a]	Number (thous.)	Percent of all households[a]
Street noise	25,754	34.8	9,412	12.7	2,864	3.9
Heavy traffic	22,475	30.4	7,484	10.1	2,590	3.5
Poor street lighting	18,024	24.4	6,552	8.9	1,042	1.4
Commercial activity	15,061	20.4	1,928	2.6	767	1.0
Crime	13,152	17.8	9,359	12.6	3,113	4.2
Airplane noise	13,140	17.8	4,319	5.8	858	1.2
Streets need repair	12,960	17.5	7,854	10.6	1,419	1.9
Trash	11,342	15.3	7,869	10.6	2,244	3.0
Roads impassable	7,880	10.6	4,311	5.8	928	1.3
Rundown housing	7,411	10.0	4,032	5.4	1,648	2.2
Odors	7,000	9.5	4,574	6.2	1,471	2.0
Abandoned structures	5,237	7.1	1,968	2.7	723	1.0

[a]Total number of households in 1976 was estimated at 74,005,000.
Source: U.S. Bureau of the Census (1978).

streets, heavy traffic, and poor street lighting. Each of these conditions was reported in the neighborhoods of an estimated 18 million or more households, which represents one in four of the Nation's households. Abandoned structures, rundown housing, and odors—the conditions whose presence was reported least often—were reported in the neighborhoods of 10 percent or fewer of the Nation's households.

These data also reveal that relatively few housholds were bothered by any one condition: no more than about one in eight households nationwide. The conditions reported as bothersome most frequently were street noise and crime (about 13 percent of all households) and least frequently were rundown housing (5.4 percent) and airplane noise and impassable roads (5.8 percent each). Crime and noisy streets were also the conditions that were often reported as extremely bothersome: about 3 million households felt that these two conditions were so bad in their neighborhoods that they wished to move because of them.

REACTIONS TO NEIGHBORHOOD CONDITIONS AS A SINGLE RESPONSE INDEX

The three responses to the surveyed conditions—presence, bothersome, and extremely bothersome—may readily be formed into a variety of different indices that summarize subjective reactions to the local environment. Such indices might consist of composite measures that combine information from all conditions into a single summary measure or measures that summarize reactions for individual conditions. The utility of indicators such as these is clear enough.

They provide unique and concise indicators of household reactions to the local environment. They may readily be used as well in relating responses to individual conditions to each other, for example, ranking them according to adverse reaction, and in relating individual responses to other variables, such as household characteristics or housing and other neighborhood characteristics.

Response indices such as these regularly serve as measures of subjective reactions to phenomena ranging from job satisfaction to crime victimization (U.S. Bureau of the Census, 1977). The recent quality of life survey by Campbell et al. (1976) provides an excellent illustration of the use of evaluation questions to provide indices for the residential environment domain. In assessing a series of local conditions and services, respondents in that survey were asked "about the way streets and roads are kept up around here. Would you say this service is very good, fairly good, neither good nor bad, not very good, or not good at all?" The five response options enabled respondents to evaluate this local condition with a set of definitive and exhaustive answers, which also provided categorical data that are both meaningful and easily interpreted without transformation or manipulation.

Responses to this question provided the following estimates: 22 percent of Americans found that their streets and roads were maintained in a very good manner, 51 percent fairly good, 7 percent neither good nor bad, 15 percent not very good, and 5 percent not good at all. A second recent national survey utilizing the same approach, evaluation question with single response dimension, reported that 7 percent of Americans found their local road and street maintenance to be excellent, 38 percent pretty good, 27 percent fair, and 26 percent poor, with 2 percent either not sure or not available (Harris, 1978, 1979).

Two single-value indicators of street and highway conditions may readily be formed from this information. With a zero assigned to the lowest ranking category of each variable—not good at all and poor—the average scores for the Nation as a whole on these two indices of street maintenance are 2.7 and 2.3. Relating these values to the response categories, one survey suggests that Americans, on the average, feel that their local streets and roads are slightly less than fairly good and the other, slightly better than pretty good. Assessments of the wide range of local conditions surveyed in the Annual Housing Survey have been dealt with in a similar manner by several researchers (Bielby, 1979; Marans, 1979).

The data presented in table 1 may be used to develop indices for each of the 12 conditions surveyed in 1976 by using the following response categories (and values):

0 Respondent indicates that the condition did not exist in the neighborhood.
1 Condition existed but was not found irritating.
2 The existing condition was found irritating.
3 The existing conditon was found to be so irritating that the household wished to relocate outside of the neighborhood.

Table 2 provides indices formed in this manner for the 12 surveyed conditions for three types of residential areas.

Several conclusions may be drawn from the scores on this index. First and perhaps foremost, they provide a single ranking of the 12 conditions. The neighborhood conditions that appear to be most bothersome from these rankings are noisy streets, heavy traffic, poorly lighted streets, and street crime. In cities, trash displaces poorly lighted streets among the conditions of greatest concern; in suburbs, airplane noise displaces street crime; and in nonmetropolitan areas, street crime is displaced by streets in need of repair. Abandoned structures appear to be the least bothersome condition in all residential locations.

Second, the values on this index are very low for all conditions regardless of residential location: all lie between 0 and 1 on a variable that could assume values as high as 3. In fact, most values are closer to the conditon-not-present category than present, but not bothersome. Such low values suggest that these conditions exist relatively infrequently and that when they do exist, they bother relatively few households.

The distribution of households across the four categories for the Nation as a whole provides some insight into why these scores are so low (table 3). The most widely reported condition, noisy streets, was reported as existing in the neighborhoods of about one-third of the Nation's households and only two other conditions, heavy traffic and poor street lighting, were reported as existing by as many as one-fourth of the Nation's households. Fewer than 1 in

Table 2. Average Values on a Four-Category Index of Responses to Neighborhood Conditions, by Residential Location

Condition	United States	Metropolitan areas			Nonmetropolitan areas
		Total	Central cities	Suburbs	
Street noise	0.51	0.55	0.61	0.49	0.45
Heavy traffic	0.44	0.46	0.53	0.40	0.40
Poor street lighting	0.35	0.33	0.26	0.40	0.37
Commercial activity	0.24	0.26	0.33	0.21	0.19
Crime	0.35	0.43	0.56	0.33	0.16
Airplane noise	0.25	0.30	0.30	0.34	0.13
Streets need repair	0.30	0.27	0.28	0.26	0.36
Trash	0.29	0.31	0.40	0.24	0.24
Roads impassable	0.18	0.17	0.19	0.16	0.19
Rundown housing	0.18	0.19	0.26	0.14	0.14
Odors	0.18	0.19	0.22	0.17	0.14
Abandoned structures	0.11	0.12	0.18	0.07	0.08

Note: Categories and their values: 0 = condition not present, 1 = condition is present, but not bothersome, 2 = condition is somewhat bothersome, 3 = condition is extremely bothersome.

Table 3. Four-Category Index of Responses to Neighborhood Conditions

(Percent)

Condition	Total	Not present	Present but not bothersome	Somewhat bothersome	Extremely bothersome
Street noise	100.0	65.2	22.1	8.8	3.9
Heavy traffic	100.0	69.6	20.3	6.6	3.5
Poor street lighting	100.0	75.6	15.5	7.4	1.4
Crime	100.0	82.2	5.1	8.4	4.2
Streets need repair	100.0	82.5	6.9	8.7	1.9
Trash	100.0	84.7	4.7	7.6	3.0
Airplane noise	100.0	82.2	11.9	4.7	1.2
Commercial activities	100.0	79.6	17.7	1.6	1.0
Rundown housing	100.0	90.0	4.6	3.2	2.2
Roads impassable	100.0	89.4	4.8	4.6	1.3
Odors	100.0	90.5	3.3	4.2	2.0
Abandoned structures	100.0	92.9	4.4	1.7	1.0

Note: Percentages may not total 100.0 because of rounding.

25 households was so bothered by a condition that it contemplated moving from the neighborhood because of it.

The extremely skewed nature of these distributions must be considered when determining what indices such as these measure. In this case, an average of 82 percent of all households reported that individual conditions did not exist in their neighborhoods; about 10 percent reported the existence of any one condition; about 6 percent were bothered by any one condition; and about 2 percent were extremely bothered. From this distribution, these indices primarily provide indicators of presence (or nonpresence) of conditions rather than subjective assessments of the nuisance levels of conditions. They do not express the concerns of respondents regarding levels of irritation produced by local conditions.

REACTIONS TO NEIGHBORHOOD CONDITIONS AS PERCEPTION AND EVALUATION MEASURES

The index introduced in the preceding section provides individual values that rank the 12 conditions in terms of respondents' reactions to them. It was shown that this particular index gives little indication of the nuisance levels of conditions. Not only that, but in this particular case, a single index masks two components of subjective responses provided by the Annual Housing Survey data, and generally considered to be independent of each other. The two independent response dimensions in this case include perceptions and evaluations, two aspects of the respondents' cognitive (memory of mental schema) image of local conditions (Rapoport, 1977).

These two reactions to environmental stimuli may best be viewed as logically

linked but representative of independent dimensions of cognition (Kaplan, 1973). The perception dimension here serves as a measure of a respondent's beliefs or recalled impressions of the presence of conditions in the neighborhood. The evaluation dimension reveals a respondent's attitudes toward conditions in terms of their nuisance levels and should be viewed as dependent upon prior experience, values, and the like, as well as upon the quantity or severity of condition actually existing in the neighborhood. The sequencing of these two dimensions suggests that the perception dimension basically reflects reality (the objective state of conditions) and that evaluations are derived of general orientation toward conditions and recalled perception of the condition (Marans and Rodgers, 1976).

In response to the general significance of these two independent reactions, the single index previously introduced is here disaggregated into two dimensions, a perception dimension and an evaluation one (the latter of which contains two criteria). Relationships among these three response measures are portrayed in figure 2. The presence measure serves as the best estimate of the actual (unmeasurable) number of households with the objective presence of conditions in their neighborhoods. The bothersome measure indicates the percentage of households bothered by a condition that they identified as existing in their neighborhoods. The extremely bothersome measure amplifies this general indicator of nuisance by reporting the number of households that were extremely bothered by a condition, that is, so bothered that they wished to move from

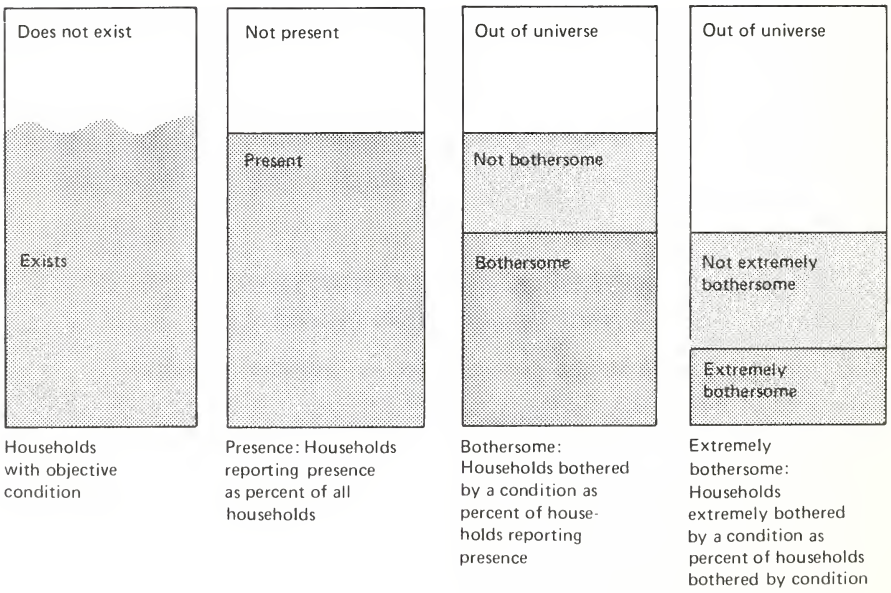

Does not exist
Exists
Households with objective condition
Not present
Present
Presence: Households reporting presence as percent of all households
Out of universe
Not bothersome
Bothersome
Bothersome: Households bothered by a condition as percent of households reporting presence
Out of universe
Not extremely bothersome
Extremely bothersome
Extremely bothersome: Households extremely bothered by a condition as percent of households bothered by condition

Figure 2. Generalized relationships between individual response measures and households with objective condition.

the neighborhood, as a percentage of those that were simply bothered. This second evaluation criterion, therefore, indicates which conditions among the bothersome ones were considered truly extreme nuisances.

The creation of a series of response measures from these data not only allows the relationships between perception and evaluation assessments to be examined but these particular measures also reflect the hierarchical structure of the questions from which responses were obtained. As previously described (figure 1), evaluation reactions were obtained only for those conditions reported by a household as present in the neighborhood. Thus, if a household reported that only one condition existed in its neighborhood, the evaluation questions were asked only for that one condition. If the household response to the first evaluation question was not bothered, then the second evaluation question was not asked. In such a case, the two evaluation questions would not be asked of the 11 conditions not reported by the respondent.

Ranking of Conditions on the Perception Measure

Values for the perception and evaluation measures are shown for three types of residential locations in table 4. Taking reports of the presence of conditions as an indicator of their actual existence, these data suggest that the most pervasive conditions nationwide in 1976 all related to streets and highways. Noisy streets, heavy traffic, poorly lighted streets, and commercial activities were all reported in the neighborhoods of between one-fourth and one-third of the Nation's households. In addition, street or traffic-related conditions were among the conditions reported as most prevalent at each of the three residential locations.

The residents of cities reported most often that their neighborhoods contained noisy streets and heavy traffic, followed by commercial activities and street crime. While suburban and nonmetropolitan residents generally reported that conditions existed less often in their neighborhoods, street-related conditions again headed their list of reported conditions. In the suburbs, noisy streets were reported more frequently than any other condition. Here, however, it was followed by poorly lighted streets, heavy traffic, and noise from aircraft, each reported by one-fifth or more of the Nation's suburban residents. In nonmetropolitan areas, the most common conditions were all street related as well: noisy streets, heavy traffic, poorly lighted streets, and streets in need of repair, all cited by at least one-fifth of nonmetropolitan area households.

Differences in the frequency of reported conditions in these three types of residential areas conform with our understanding of actual differences between these areas, for example, reported crime rates, residential densities, heterogeneity of land use at different scales, and levels of public services. In general, however, the rankings produced for the three residential areas agree rather closely, primarily because conditions associated with streets and highways, which are such an integral aspect of residential areas regardless of location, head the rankings in each location. Correlations (Spearman's rank-order coefficients)

Table 4. Perception and Evaluation Response Measures to Neighborhood Conditions, by Residential Location

Condition	United States	Metropolitan areas			Nonmetropolitan areas
		Total	Central cities	Suburbs	
	Perception measure				
Street noise	34.8	36.1	39.9	33.0	31.9
Heavy traffic	30.4	30.8	35.4	27.0	29.5
Poor street lighting	24.4	22.1	15.1	28.0	29.1
Commercial activity	20.4	22.3	27.9	17.4	16.4
Crime	17.8	22.0	27.7	17.3	8.7
Airplane noise	17.8	21.3	21.0	21.5	10.2
Streets need repair	17.5	15.5	15.7	15.3	21.9
Trash	15.3	16.0	19.8	12.9	13.9
Roads impassable	10.6	10.2	11.1	9.4	11.6
Rundown housing	10.0	10.6	13.9	7.9	8.7
Odors	9.5	10.2	11.4	9.2	7.9
Abandoned structures	7.1	7.5	11.0	4.6	6.2
	Bothersome measure				
Street noise	36.5	38.5	39.0	38.0	31.8
Heavy traffic	33.3	35.9	34.8	37.1	27.5
Poor street lighting	36.4	43.5	55.0	38.3	24.8
Commercial activity	12.8	13.2	12.5	14.1	11.7
Crime	71.2	71.9	71.6	72.3	67.0
Airplane noise	32.9	34.5	35.6	33.6	25.6
Streets need repair	60.6	61.9	62.5	61.5	58.6
Trash	69.4	72.4	72.8	71.9	62.0
Roads impassable	54.7	55.8	57.0	54.7	52.6
Rundown housing	54.4	57.9	58.2	57.4	45.4
Odors	65.3	66.6	65.8	67.4	61.9
Abandoned structures	37.6	42.3	44.0	39.0	25.4
	Extremely bothersome measure				
Street noise	30.4	32.2	35.5	28.8	25.2
Heavy traffic	34.6	35.9	39.5	32.2	30.8
Poor street lighting	15.9	17.0	25.6	11.4	12.8
Commercial activity	39.8	41.5	43.8	38.8	34.4
Crime	33.3	35.4	42.2	26.4	21.0
Airplane noise	19.9	20.9	23.0	19.0	13.8
Streets need repair	18.1	20.6	24.9	16.9	13.9
Trash	28.5	31.8	38.0	23.9	14.0
Roads impassable	21.5	23.3	26.7	19.8	17.9
Rundown housing	40.9	43.5	49.4	34.9	32.0
Odors	32.2	34.5	39.5	29.5	25.2
Abandoned structures	36.7	41.0	45.4	30.9	19.0

Note: Perception measure scores are percentages of households reporting the presence of conditions. Bothersome measure scores are the number of households reporting conditions to be bothersome as a percentage of those reporting that the condition existed in the neighborhood. Extremely bothersome measure scores represent the number of households that were so bothered that they wished to leave the neighborhood as a percentage of households bothered by the condition.

between the presence rankings in the three locations reflect the general agreement among residents reporting conditions in their neighborhoods in the three major types of residential areas

$$r_s = \begin{cases} 0.80 & \text{for city-suburban residents} \\ 0.82 & \text{for suburban-nonmetropolitan residents} \\ 0.67 & \text{for city-nonmetropolitan residents} \end{cases}$$

Ranking of Conditions in the Evaluation Measures

Rankings of the frequency with which households reported being bothered by conditions again demonstrate a high level of conformity between households located in each of the three residential areas. Correlations between the three bothersome rankings are

$$r_s = \begin{cases} 0.97 & \text{for city-suburban residents} \\ 0.93 & \text{for suburban-nonmetropolitan residents} \\ 0.85 & \text{for city-nonmetropolitan residents} \end{cases}$$

The same four conditions, street crime, trash, odors, streets needing repair, were reported most frequently as being bothersome at each residential location. The presence of retail establishments, industrial facilities, or other commercial activities bothered persons least often.

Rankings on the extremely bothersome measure were also similar across the three types of residential areas: correlations between the rankings in each area were once again high:

$$r_s = \begin{cases} 0.87 & \text{for city-suburban residents} \\ 0.94 & \text{for suburban-nonmetropolitan residents} \\ 0.77 & \text{for city-nonmetropolitan residents} \end{cases}$$

Differences do exist, however, among the conditions ranked at the top of the most severe reaction to conditions. Rundown housing and commercial activities ranked at or near the top of this measure regardless of location. In cities, however, crime and abandoned structures were also found at the top of the list; in suburbs, heavy traffic and abandoned structures were added to rundown housing and commercial activities; and in nonmetropolitan areas, heavy traffic and odors.

Relationships Between Perception and Evaluation Measures

In forming single response indices that combine information on both the perception and evaluation of conditions, one assumes that these two types of reactions are additive; that is, those conditions perceived as existing most often are also evaluated as bothersome most frequently or most bothersome. If in

fact these two different types of reactions are complementary, the order of conditions on the presence, bothersome, and extremely bothersome measures should be similar. In other words, those conditions relating to streets, such as street noise, high traffic volumes, and inadequate street lighting, which are the conditions most frequently reported as existing, should also be evaluated as bothersome and extremely bothersome most often. Such a pattern of common rankings on the three measures does not emerge, however.

The ordering of conditions on the three measures follows not one simple pattern but four. The first of the four patterns is created by conditions that households reported as existing relatively seldom, rated higher on the bothersome measure and still higher on the extremely bothersome measure. Two conditions, rundown housing and abandoned structures, followed this pattern at each of the three residential locations. The opposite pattern, conditions ranked highest on the presence measure and lowest on the extremely bothersome measure, held for airplane noise in all three residential locations and for poorly lighted streets in suburbs and nonmetropolitan areas.

The other two patterns of shifts across the three measures were comprised of conditions that were either lower on the bothersome measure than on the other two or higher on the bothersome measure than on the others. The first of these two patterns was followed by five conditions in all three residential locations: street crime, trash, streets needing repair, odors, and impassable roads. Poorly lighted streets followed the same pattern in cities. Conditions following the reverse of this pattern, that is, ranked lower on the bothersome measure than on either of the other two, included street noise, heavy traffic, and commercial activities.

These variations in the rankings of conditions on the three response measures reveal important differences about the ways in which people perceive and evaluate conditions in their residential environments. Rundown housing and abandoned structures, for instance, which probably serve better than any other conditions as indicators of severely deteriorated residential conditions, occur relatively infrequently and are found near the bottom of the presence measure in all three residential locations. In all three locations, however, their rankings rise relative to other conditions on the two evaluation measures, rating highest on the extremely bothersome measure.

Airplane noise represents a condition that followed just the opposite pattern of shifts across the three response measures. Whereas its existence was reported a relatively large number of times (especially in cities and suburbs), it was found bothersome less often than any other condition in cities and less often than all conditions except streets needing repair and poorly lighted streets in suburbs and nonmetropolitan areas. It represents, therefore, one of the conditions that was widely perceived as existing but not of very much concern relative to a large number of other local conditions.

Commercial activities was one of several conditions ranked relatively lower on

the bothersome measure than on either of the other two. In each residential location, commercial activities were reported an average number of times relative to the other surveyed conditions. However, the proportion of households evaluating the presence of such activities as bothersome was lower than for other conditions: it was at the bottom of the bothersome measure at each location. Nonetheless, of the households that were bothered by the presence of commercial activities, a very large proportion were extremely bothered. It ranked at the top of the extremely bothersome measure at all locations except cities, where rundown housing and abandoned structures were reported to be extremely bothersome more often, suggesting that the condition was not generally a nuisance but can be among the most bothersome of all local irritants.

Street crime represents one of the conditions with a pattern of shifts the opposite of commercial activities. While its position varied somewhat on the presence measure (reported more frequently in cities than suburbs or nonmetropolitan areas), street crime was found bothersome more frequently than any other condition in suburbs and nonmetropolitan areas and second only to trash in cities. In each location, however, several other conditions were reported as extremely bothersome more often, for example, rundown housing and commercial activities.

The shifts in rank that occur between each of these three measures strongly suggest that the perception and evaluation response dimensions are independent of each other and therefore ought to be used individually because each represents a fundamentally different type of response to conditions on the neighborhood scale. Each response also provides additional information in attempting to determine which conditions are of greatest concern to local residents. The fact that a nearly identical pattern of rankings occurs among measures in residential environments that differ as much as those of cities, suburbs, and nonmetropolitan areas emphasizes the fundamental nature of these subjective responses.

CONCLUSIONS

The reactions of households to 12 neighborhood conditions were analyzed from a number of perspectives in order to isolate those conditions that respondents felt were the least desirable and that therefore are associated with the Nation's lowest quality residential environments. Initially the full set of responses was formed into single response indices for each of the 12 surveyed conditions. Discussion of this measure's utility pointed out the distinct advantage it provides in producing a single response value for each condition. It was noted, however, that such a single response measure with Annual Housing Survey data is overwhelmingly influenced by the perception of the presence of conditions and little influenced by the evaluations of their nuisance levels. It therefore serves poorly as a measure of dissatisfaction.

In response to this limitation and reflecting the fact that the single response index combined information from two unique dimensions—perception and

evaluation—this measure was disaggregated into component measures to represent these two independent dimensions. The first of the three response measures formed through disaggregation, the presence dimension, reported households' capacity to recall the presence of conditions in their neighborhoods. The evaluation dimension consisted of two response criteria—a bothersome measure, which reported whether households were bothered by existing conditions, and an extremely bothersome measure, which amplified the basic bothersome measure by indicating which bothersome conditions were so bothersome that the household wished to move from the area because of them.

Each of the response measures provided consistent results among residents in three different types of residential location (cities, suburbs, and nonmetropolitan areas). The variations that did occur between these three types of areas, such as the higher ranking given to crime on the presence measure in cities, reflect characteristics relatively unique to each of the three places of residence. The stability of rankings on the three disaggregated response measures among these different residential settings suggests that comparisons between the measures may be made at the national level without significant loss of information.

Table 5 presents rankings of the conditions on each of the four response measures plus a list of the conditions ordered by their relationship to overall neighborhood dissatisfaction. From the previous discussion, it is not surprising to discover that the national rankings of conditions on the composite index are extremely similar to those on the presence dimension alone. The high correlation between these two rankings, single response index-perception measure, r_s = 0.91, reflects the fact that only five conditions differ in their rank order between the two measures and only commercial activity differs by as many as four positions. The level of congruence between the single response index and the other two disaggregated measures is considerably lower; for single response index-bothersome measure, $r_s = -0.15$, and for single response index-extremely bothersome measure, $r_s = -0.38$, reflecting again the basic independence of the perception and evaluation dimensions.

The lack of congruence between the disaggregated response measures suggests that the perception and evaluation dimensions must both be considered in answering the basic question: Which neighborhood conditions concern Americans the most and therefore are in greatest need of improvement if we are to achieve the "suitable environment" portion of the national housing goal? First and foremost, these results indicate that the most prevalent conditions, regardless of residential location, were all street-related: noisy streets, heavy volumes of street traffic, poorly lighted streets, and the presence of various types of commercial activities. Zoning and other forms of land use control have been utilized for a considerable period of time as a means of reducing levels of traffic and street noise and for limiting the presence of commercial activity in residential areas. Nonetheless, the basic functional relationship between places of residence and the transportation system (especially highways) requires some mixing of these different land uses, so we can therefore continue to expect street-related conditions to be widely reported in residential areas.

Table 5. Rankings of Neighborhood Conditions on Response Measures and Relationship to Neighborhood Dissatisfaction

Single response index	Multiple response dimensions			Overall correlation with neighborhood dissatisfaction
	Presence	Bothersome	Extremely bothersome	
Street noise	Street noise	Crime	Rundown housing	Rundown housing
Heavy traffic	Heavy traffic	Trash	Commercial activity	Abandoned structures
Poor street lighting	Poor street lighting	Odors	Abandoned structures	Crime
Crime	Commercial activity	Streets need repair	Heavy traffic	Trash
Streets need repair	Crime	Roads impassable	Crime	Commercial activities
Trash	Airplane noise	Rundown housing	Odors	Street noise
Airplane noise	Streets need repair	Abandoned structures	Street noise	Streets need repair
Commercial activity	Trash	Street noise	Trash	Poor street lighting
Roads impassable	Roads impassable	Poor street lighting	Roads impassable	Heavy traffic
Rundown housing	Rundown housing	Heavy traffic	Airplane noise	Odors
Odors	Odors	Airplane noise	Streets need repair	Roads impassable
Abandoned structures	Abandoned structures	Commercial activity	Poor street lighting	Airplane noise

Although widely reported as existing, street-related conditions were not judged to be bothersome nearly as often as other conditions. In fact, they ranked at the bottom of the bothersome measure along with airplane noise. Conditions more frequently reported as bothersome reflect more serious concerns about safety and several other aspects of the local environment as an integral component of housing. Street crime, reported in the neighborhoods of nearly one in six of the Nation's households (17.8 percent) was found bothersome by 7 out of 10 households (71.2 percent) reporting its existence. One-third of this latter group of households also felt that crime in their neighborhoods was bad enough to want to move because of it.

The second evaluation measure amplifies the bothersome rankings by differentiating between those households that were simply bothered by conditions and those that found them to be so bad that they wished to move to avoid them. Conditions at the top of this dimension must be considered the most undesirable of the bothersome conditons. Rundown housing, commercial activities, abandoned structures, heavy traffic, and crime were all reported as so bothersome that one-third or more of the households bothered by them also wished to leave their neighborhoods because of them.

A perspective on these assessments of neighborhood conditions from the household's global view of the neighborhood may be obtained by relating the 12 conditions to overall neighborhood dissatisfaction. The ranking of conditions on this measure (the final column of table 5) shows that households reporting the presence of rundown housing, abandoned structures, crime, and trash found their neighborhoods to be less satisfactory places to live than those who reported the presence of conditions such as noise from aircraft or impassable roads.

A comparison of this global indicator of neighborhood dissatisfaction with the three disaggregated response measures confirms the importance of the two evaluation dimensions as indicators of nuisance level. The relationship between conditions ranked by their association with overall neighborhood satisfaction and conditions on the disaggregated response measures becomes increasingly more congruent with each response measure:

$$r_s = \begin{cases} -0.20 & \text{for overall satisfaction-presence} \\ -0.27 & \text{for overall dissatisfaction-bothersome} \\ 0.61 & \text{for overall dissatisfaction-extremely bothersome} \end{cases}$$

The negative relationship between the global ranking and the presence measure emphasizes the fact that not all of these conditions are perceived as particularly significant in the evaluation of neighborhood quality. Some conditions, however, rank high on both evaluation measures and the dissatisfaction measure and must therefore be considered as deserving the most attention in attempting to improve the quality of the Nation's neighborhoods. Conditions in this category include rundown housing, abandoned structures, crime, trash, and commercial activities.

In conclusion, these findings suggest that consideration must be given to differentiating between the perception and evaluation of conditions in producing subjective indicators of neighborhood quality for a number of very specific and important reasons. Item-by-item levels of perceived presence and dissatisfaction are not necessarily closely related. In fact, inverse relationships were shown to exist between the rankings of conditions on the perception and evaluation dimensions in each of the three major types of residential areas. In addition, the relatively low rates at which persons reported that conditions were objectionable calls into question the use of the two dimensions to create single indices for each condition or create a composite index that would incorporate information from all conditions.

By disaggregating responses into their perception and evaluation dimensions, it was demonstrated that street-related conditions (e.g., noisy streets, heavy traffic, and poorly lighted streets), while perceived as existing in large numbers of neighborhoods, were not found to be as bothersome as various aspects of the local built environment (e.g., rundown housing, abandoned structures, and commercial activities) and personal safety. While these latter conditions were reported in relatively few areas and represent problems whose solutions must be sought from widely diverging realms, they are the most severe problems in the eyes of residents. They should, therefore, be given careful attention in considering the quality of the Nation's neighborhoods and in measuring progress toward achieving the suitable environment portion of the Nation's housing goal.

REFERENCES

Andrews, Frank M., and Stephen B. Withey. 1976. *Social Indicators of Well-Being: Americans' Perceptions of Life Quality*. New York: Plenum Press.

Antunes, George E., and John P. Plumlee. 1977. "The Distribution of an Urban Public Service." *Urban Affairs Quarterly* 12: 313-332.

Bauer, Raymond A. 1967. *Social Indicators*. Cambridge, Mass.: M.I.T. Press.

Bell, Daniel. 1969. "The Idea of a Social Report." *The Public Interest* 15: 72-84.

Berry, Brian J. L., et al. 1974. *Land Use, Urban Form and Evironmental Quality*. Research Paper no. 155. Chicago: Department of Geography, University of Chicago.

____. 1976. *Social Burdens of Environmental Pollution*. Cambridge, Mass.: Ballinger.

Bielby, William T. 1979. *Evaluating Measures of Neighborhood Quality in the Annual Housing Survey*. Annual Housing Survey Studies, no. 2, U.S. Department of Housing and Urban Development. Washington, D.C.: U.S. Government Printing Office.

Campbell, Angus, Philip E. Converse, and Willard L. Rodgers. 1976. *The Quality of American Life: Perceptions, Evaluation, and Satisfaction*. New York: Russell Sage Foundation.

Dahmann, Donald C. 1979. "North-South Differences in Housing and the

Housing and Community Development Act of 1977." *Journal of the American Real Estate and Urban Economics Association* 7: 230-242.

Frieden, Bernard J., and Arthur P. Solomon. 1977. *The Nation's Housing: 1975-1985.* Cambridge, Mass.: Joint Center for Urban Studies of M.I.T. and Harvard.

Gans, Herbert J. 1962. *The Urban Villagers.* New York: Free Press.

_____. 1967. *The Levittowners.* New York: Random House.

Goldblatt, Harold. 1977. "The Quality of Our Streets and Neighborhoods." *HUD Challenge* March: 4-8.

_____. 1979. *A Survey of Citizen Views and Concerns About Urban Life.* Washington, D.C.: U.S. Government Printing Office.

Hirschman, Albert O. 1970. *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States.* Cambridge, Mass.: Harvard University Press.

Hoch, Irving. 1972. "Urban Scale and Environmental Quality." In Ronald G. Ridker, ed., *Population Resources and the Environment,* Research Reports, vol. 3. U.S. Commission on Population Growth and the American Future. Washington, D.C.: U.S. Government Printing Office.

Hunter, Albert. 1974. *Symbolic Communities: The Persistence and Change of Chicago's Local Communities.* Chicago: University of Chicago Press.

Isler, Morton L. 1970. *Thinking About Housing: A Policy Research Agenda.* Washington, D.C.: The Urban Institute.

Kaplan, Stephen. 1973. "Cognitive Maps in Perception and Thought." In Roger M. Downs and David Stea, eds., *Image and Environment: Cognitive Mapping and Spatial Behavior,* pp. 63-78. Chicago: Aldine.

Lansing, John B., and Robert W. Marans. 1969. "Evaluation of Neighborhood Quality." *Journal of the American Institute of Planners* 35: 195-199.

Louis Harris and Associates, Inc. 1978. *1978 HUD Survey on the Quality of Community Life.* Washington, D.C.: U.S. Government Printing Office.

Marans, Robert W. 1979. *The Determinants of Neighborhood Quality: An Analysis of the 1976 Annual Housing Survey.* Annual Housing Survey Studies, no. 3, U.S. Department of Housing and Urban Development. Washington, D.C.: U.S. Government Printing Office.

Marans, Robert, and Willard L. Rodgers. 1976. "The Residential Environment." In Angus Campbell et al., *The Quality of American Life.* New York: Russell Sage Foundation.

Michelson, William. 1966. "An Empirical Analysis of Urban Environmental Preference." *Journal of the American Institute of Planners* 32: 355-360.

Newman, Sandra J., and Greg J. Duncan. 1979. "Residential Problems, Dissatisfaction, and Mobility." *Journal of the American Planning Association* 45: 154-166.

Olson, Mancur, Jr. 1969. "Social Indicators and Social Accounts." *Socio-Economic Planning Sciences* 2: 335-346.

Park, Robert E. 1916. "The City: Suggestions for the Investigation of Human Behavior in the Urban Environment." *American Journal of Sociology* 20: 577-612. Also in Robert E. Park et al. 1925. *The City,* pp. 1-46. Chicago: University of Chicago Press.

Rapoport, Amos. 1977. *Human Aspects of Urban Forms: Towards a Man-Environment Approach to Urban Form and Design.* Oxford: Pergamon Press.

Russell Sage Foundation. 1909-1914. *The Pittsburgh Survey,* 6 vols. New York: Russell Sage Foundation.

Schmid, James A. 1974. "The Environmental Impact of Urbanization." In Ian R. Manners and Marvin W. Midesell, eds., *Perspectives on Environment.* Washington, D.C.: Association of American Geographers.

Sheldon, Eleanor B., and Wilbert Moore, eds. 1968. *Indicators of Social Change: Concepts and Measurement.* New York: Russell Sage Foundation.

Stein, Maurice R. 1960. *The Eclipse of Community.* Princeton, N.J.: Princeton University Press.

Suttles, Gerald D. 1968. *The Social Order of the Slum: Ethnicity and Territory in the Inner City.* Chicago: University of Chicago Press.

U.S. Bureau of the Census. 1977. *Social Indicators, 1976.* Washington, D.C.: U.S. Government Printing Office.

____. 1978. "Indicators of Housing and Neighborhood Quality." *Current Housing Reports,* series H-150-76, part B. Washington, D.C.: U.S. Government Printing Office.

U.S. Council on Environmental Quality. 1970 and thereafter. *Annual Report.* Washington, D.C.: U.S. Government Printing Office.

U.S. Department of Health, Education, and Welfare. 1970. *Toward a Social Report.* Ann Arbor, Mich.: University of Michigan Press.

U.S. Department of Housing and Urban Development. 1979a. *A Guide to the Annual Housing Survey.* Washington, D.C.: U.S. Government Printing Office.

____. 1979b. *The Tenth Annual Report on the National Housing Goal.* Washington, D.C.: U.S. Government Printing Office.

U.S. National Commission on Neighborhoods. 1979. *People, Building Neighborhoods: Final Report and Case Study Appendices,* vols. 1 and 2. Washington, D.C.: U.S. Government Printing Office.

U.S. Office of Management and Budget. 1976. *Standard Metropolitan Statistical Areas: 1975,* rev. ed. Washington, D.C.: U.S. Government Printing Office.

U.S. President's Research Committee on Social Trends. 1933. *Recent Social Trends in the United States.* New York: McGraw-Hill.

Wolpert, Julian. 1966. "Migration As An Adjustment to Environmental Stress." *Journal of Social Issues* 22(4): 92-102.

Zehner, Robert B. 1971. "Neighborhood and Community Satisfaction in New Towns and Less Planned Suburbs." *Journal of the American Institute of Planners* 37: 379-385.

Can We Have Confidence in Confidence? Revisited

Tom W. Smith
National Opinion Research Center
University of Chicago

For 5 years (1972-1977), the General Social Survey (GSS) at the National Opinion Research Center (NORC) and Louis Harris and Associates have been asking parallel series on confidence in major national institutions. Comparison of NORC and Harris findings indicated that discrepancies frequently existed between the two series. These differences caught the attention of the GSS staff and other scholars (e.g., Ladd, 1976-1977; Santi, 1978; Turner and Krauss, 1978) and led to an evaluation of these differences (Smith, 1977). To examine the reason for the differences more fully and in general to understand better the meaning and nature of the confidence questions, a number of methodological experiments were conducted on the 1978 GSS. This chapter updates the initial evaluation of differences between the NORC and Harris series and analyzes the confidence experiments. First, a detailed examination is made of sample population, question wording, format, placement, and related matters. The question here is simply how similar the two series are in form. Second, the marginals generated by the two series are compared. Inspection is made of the marginal differences at the same time points, of the trends of each series, of overall confidence averages, and of rank ordering. Third, the findings of the methodological experiments and other evidence are examined and these findings are related to the differences between NORC and Harris findings. Finally, the use of the confidence items as social indicators is considered.

COMPARISON OF FORMS

Altogether Harris has asked the confidence question 22 times and NORC 6 times between 1966 and 1977.[1]

[1] Both NORC and Harris have subsequently extended their series. This analysis covers surveys through January 1977 for Harris and March 1977 for NORC. For NORC, see Davis, Smith, and Stephenson, 1980.

Harris 1574	February 1966	Harris 2430	September 1974
Harris 1702	January 1967	Harris 2434	September 1974
Harris 2131	August 1971	Harris 2515	March 1975
Harris 2219	May 1972	GSS 1975	March 1975
Harris 2236	October 1972	Harris 7581	April 1975
Harris 2251	January 1973	Harris 7585	August 1975
GSS 1973	March 1973	Harris 2521	February 1976
Harris 2319	March 1973	Harris 7681	March 1976
Harris 2343	September 1973	GSS 1976	March 1976
Harris 2354	December 1973	Harris 7684	June 1976
NORC 4179	January 1974	Harris 2628	July 1976
Harris 7482	February 1974	Harris 2630	July 1976
GSS 1974	March 1974	Harris 7690	January 1977
Harris 7487	August 1974	GSS 1977	March 1977

All NORC surveys cover the noninstitutionalized population of the continental United States, 18 years and over. This universe is also covered by Harris 1574, 2131, 2219, 2251, 2343, 2354, 7482, 7487, 2430, 2434, 2515, 2521, 2628, 7581, 7585, 7681, and 7690. Four Harris surveys were limited to electoral participators. Harris 1702 was restricted to those 21 and older, who voted in 1960, 1962, 1964, or 1966 or were registered to vote. Harris 2236 covered those 18 and older, who voted in 1968 or 1970, were registered to vote, or were under 25 years old. Harris 7684 and 2630 were limited to those who were either over 30 and voted in 1972 or 1974 or 18-29 and registered to vote or planning to register. Finally, Harris 2319 sampled those 16 and older who worked at least 35 hours per week. In addition, Harris 1574 included a sample of teenagers along with its adult sample, and Harris 2343 had a supplementary sample of State and local officials. Two special surveys not discussed here include Harris 1577S, a student sample that included a related confidence in occupation item, and Harris 2522, a sample of South Carolina.

The original use of the item by Harris in 1966 went as follows:

> As far as the people running (*read list*) are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them?

This wording was repeated on Harris 1702, 2131, 2236, 2251, 2319, and 2354 (except for variations in the interviewer instructions; see table 1). Since then, Harris has used six variations of this original wording (table 1). The GSS version was based on the original Harris wording and has remained unchanged.

> I am going to name some institutions in this country. As far as the *people running* these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them?

Table 1. Wording Variations on Harris Confidence Questions

1. As far as the people running (*read list*) are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them? (*Record below for each item on list.*)

   2131, 2354
   1574, 1702 (*Rotate asking order and record below*)
   2236 (*Record below and continue with list*)
   2251, 2319 (*Read first item on list*) . . . (*Record below and continue with list.*)

2. Now let me ask you how much confidence you have in people who are in charge of managing (*read list*)—a great deal of confidence, only some, or hardly any confidence? (*Record below for each item on list.*)

   2219

3. As far as the people in charge of running (*read list*) are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them? (*Record below for each item on list.*)

   2343, 2430, 2434, 7487, 7581, 7681
   7482 Now, . . . (*Read first item on list*) . . . (*Record below and continue with list.*)

4. How much confidence do you have in the people running (*read first item on list*)—a great deal of confidence, only some, or hardly any confidence at all? (*Record below and continue with list.*)

   2515, 2628

5. As far as people in charge of running (*read list*) are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them? (*Record below for each item on list.*)

   7690, 7585, 2521

6. How much confidence do you have in (*read list*)—a great deal, only some, or hardly any? (*Record below for each item.*)

   7684

7. As far as the people running (*read list*) are concerned, do you have a great deal of confidence in them, only some, or hardly any confidence in them at all? (*Record below for each item on list.*)

   2630

Note: Italics indicate instructions to interviewer.

The NORC amalgam version (NORC 4187) is a variation of the GSS wording.

> Now I am going to name some institutions in this country. As far as the *people running* these institutions are concerned, would you say you have a great deal of confidence in them, only some confidence, or hardly any confidence at all in them?
> *Read each item; code one for each. Repeat the question or categories as necessary.*

The various Harris and NORC versions differ in many particulars, although several basic similarities exist. The Harris versions all start off by inquiring about a list of institutions embedded into the question. At the appropriate place the interviewer is instructed to *read list* or some similar command and

then at the end of the questions an instruction such as *record below and continue with list* appears. The GSS and NORC amalgam versions on the other hand preface the inquiry about confidence with an introductory sentence and replace the insertion of enumerated institutions with the general phrase "these institutions." On the GSS surveys and NORC 4179 the interviewer is then instructed to *read each item: code one for each*. On the 1973, 1974, and 1977 GSS and NORC 4179 the additional instruction *repeat the question or categories as necessary* appears. In 1975 and 1976, necessary repetitions were built into the question at the first, second, sixth, and eleventh items, as follows:

A. First, how much confidence do you have in the people running . . .
B. How about people running . . .
C.
D.
E.
F. How about the people running . . .
G.
H.
I.
J.
K. How about the people running . . .

Also, at this point on the 1974-1977 GSS surveys, the interviewer handed the respondent a card with the response categories listed vertically in the order they were read. On GSS 1973 and NORC 4179, no card was used. No card is used on any of the Harris surveys.

Other differences also exist on the wordings of various parts of the question. The original Harris item and some later versions (wordings 1, 4, and 7 in table 1) ask about the people running the enumerated institutions. Later most surveys refer to the "people in charge of running" the listed institutions (wordings 3 and 5). One variation refers to "people who are in charge of managing" (wording 2), and another variation mentions only the institutions themselves (wording 6). The GSS and NORC amalgam versions follow the original Harris allusion to the people running and this phrase is emphasized.

The response categories on the Harris surveys differ in several slight ways, but they generally consist of *a great deal of confidence, only some confidence,* and *hardly any confidence at all in them* (see table 1 for the variations). In Harris 1574, 1702, 2131, 2236, and 2319, however, an unsolicited response of *none* was precoded in the questionnaire for those who volunteered this answer. The GSS always has the following response categories: a great deal of confidence, only some confidence, or hardly any confidence at all in them. NORC 4179 varies by inserting "in them" after a great deal of confidence.

Differences also exist in the wording, order, and number of enumerated institutions. Table 2 compares the descriptors used in the surveys. (GSS covered

institutions only.) The institutions are referred to by the following acronyms:

CONARMY	Military
CONBUS	Major companies
CONCLERG	Organized religion
CONEDUC	Education
CONFED	Executive branch of the Federal Government
CONFINAN	Banks and financial institutions
CONJUDGE	U.S. Supreme Court
CONLABOR	Organized labor
CONLEGIS	Congress
CONMEDIC	Medicine
CONPRESS	Press
CONSCI	Scientific community
CONTV	Television

Of the 13 institutions, only 2 (medicine and organized religion) remain unchanged across all surveys. On two items (education and television) Harris alters their original descriptors to cover more specific institutions (higher education and television news), while the GSS continues to cover the originally chosen institutions (although television is replaced with TV). On two other items referring to the Armed Forces and the judicial system, items vary only by whether the article *the* is used. This pattern also occurs on the executive branch item along with a variant, *the Federal Government*, in Harris 2319; on the science question along with the variant *science* in Harris 7581 and 2630; and on the press question along with the variant *television and the press* in Harris 2319 and *the newspapers* in Harris 2219. The legislative question has a single variant, broken down into separate items on the Senate and the House of Representatives in Harris 2343, as does the business question, which appears as *large business corporations* in Harris 2219. Labor also has two versions, the standard *organized labor*, and *labor unions* on Harris 2219, 2515, 7581, and 2628. Showing the greatest variation is the banking item, which appears in five different forms.

Most variations occur between Harris and GSS or within the Harris series. The GSS varies only by the use of the article *the* in the science item in 1975 and 1976 and its omission in other years.

In addition to the description changes, there were numerous alterations in the order in which the items were asked (table 3). What might be called the standard order occurred on Harris 1574, 1702, 2131, 2236, and 2251. GSS 1973 and 1974 retained this order but omitted items 9, 13, and 15 from the Harris list. GSS 1977 also followed the original order through item 8 and then asked CONFINAN in the last position. All other surveys follow different patterns, and except for GSS 1975 and 1976 and Harris 7585 and 7690, each is unique. The ordering changes are not random, however, and certain suborderings hold up through most but not all surveys. For example, confidence in

Table 2. Item Descriptors, by Survey

Survey	CONBUS	CONCLERG	CONEDUC	CONFED	CONPRESS	CONMEDIC
Harris 1574	Major companies	Organized religion	Education	Executive branch of Federal Government	The press	Medicine
Harris 1702	Major companies	Organized religion	Education	Executive branch of the Government	The press	Medicine
Harris 2131	Major companies	Organized religion	Education	Executive branch of Government	The press	Medicine
Harris 2219	Large business corporations	(X)	Colleges and universities	The Federal Government	The newspapers	(X)
Harris 2236	Major companies	Organized religion	Education	Executive branch of Government	The press	Medicine
Harris 2251	Major companies	Organized religion	Education	The executive branch of the Federal Government	The press	Medicine
Harris 2319	Major companies	Organized religion	(X)	The Federal Government	Television and the press	Medicine
GSS 1973	Major companies	Organized religion	Education	The executive branch of the Federal Government	Press	Medicine
Harris 2343	Major companies	Organized religion	Higher educational institutions (colleges, universities, etc.)	The executive branch of the Federal Government	The press	Medicine
Harris 2354	Major companies	Organized religion	Higher education (colleges and universities)	Executive branch of Government	The press	Medicine
NORC 4179	Major companies	(X)	(X)	Executive branch of the Federal Government	The press	(X)

Harris 7482	Major companies	(X)	(X)	Executive branch of the Federal Government	(X)	Medicine
GSS 1974	Major companies	Organized religion	Education	Executive branch of the Federal Government	Press	Medicine
Harris 7487	Major companies	Organized religion	Higher educational institutions (colleges, universities, etc.)	The executive branch of the Federal Government	The press	Medicine
Harris 2430	Major companies	(X)	(X)	The executive branch of the Federal Government	The press	Medicine
Harris 2434	Major companies	Organized religion	Higher educational institutions (colleges, universities, etc.)	The executive branch of the Federal Government	The press	Medicine
Harris 2515	Major companies	(X)	Higher education	(X)	The press	(X)
GSS 1975	Major companies	Organized religion	Education	The executive branch of the Federal Government	Press	Medicine
Harris 7581	Major companies	Organized religion	Higher education (colleges and universities)	The executive branch of the Federal Government	The press	Medicine
Harris 7585	Major companies	Organized religion	Higher educational institutions (colleges, universities, etc.)	The executive branch of the Federal Government	The press	Medicine
Harris 2521	Major companies	Organized religion	Higher educational institutions (colleges universities, etc.)	The executive branch of the Federal Government	The press	Medicine

(See footnote on p. 129.)

Table 2. Item Descriptors, by Survey—Continued

Survey	CONBUS	CONCLERG	CONEDUC	CONFED	CONPRESS	CONMEDIC
Harris 7681	Major companies	(X)	(X)	The executive branch of the Federal Government	The press	(X)
GSS 1976	Major companies	Organized religion	Education	The executive branch of the Federal Government	Press	Medicine
Harris 7684	(X)	(X)	(X)	The executive branch of the Federal Government	(X)	(X)
Harris 2628	Major companies	(X)	Higher education	(X)	The press	(X)
Harris 2630	Major companies	(X)	(X)	Executive branch of the Federal Government	The press	Medicine
Harris 7690	Major companies	Organized religion	Higher educational institutions (colleges, universities, etc.)	The executive branch of the Federal Government	The press	Medicine
GSS 1977	Major companies	Organized religion	Education	The executive branch of the Federal Government	Press	Medicine

Survey	CONTV	CONFINAN	CONJUDGE	CONSCI	CONLEGIS	CONARMY	CONLABOR
Harris 1574	Television	Banks and financial institutions	The U.S. Supreme Court	The scientific community	Congress	The military	Organized labor
Harris 1702	Television	Banks and financial institutions	The U.S. Supreme Court	The scientific community	Congress	The military	Organized labor

Harris 2131	Television	Banks and financial institutions	The U.S. Supreme Court	The scientific community	Congress	The military	Organized labor
Harris 2219	Television networks	Full service banks	(X)	(X)	(X)	(X)	Labor unions
Harris 2236	Television	Banks and financial institutions	The U.S. Supreme Court	The scientific community	Congress	The military	Organized labor
Harris 2251	Television	Banks and financial institutions	The U.S. Supreme Court	The scientific community	Congress	The military	Organized labor
Harris 2319	See CONPRESS	Banks and other financial institutions	(X)	(X)	(X)	(X)	Organized labor
GSS 1973	TV	(X)	U.S. Supreme Court	Scientific community	Congress	Military	Organized labor
Harris 2343	Television news	(X)	The U.S. Supreme Court	(X)	The U.S. House of Representatives/The U.S. Senate	The military	Organized labor
Harris 2354	TV news	Banking	(X)	The scientific community	Congress	(X)	Organized labor
NORC 4179	(X)	(X)	The U.S. Supreme Court	(X)	Congress	(X)	(X)
Harris 7482	TV news	(X)	(X)	(X)	(X)	(X)	(X)
GSS 1974	TV	(X)	U.S. Supreme Court	Scientific community	Congress	Military	Organized labor
Harris 7487	Television news	(X)	The U.S. Supreme Court	(X)	Congress	The military	Organized labor

(See footnote on p. 129.)

Table 2. Item Descriptors, by Survey—Continued

Survey	CONTV	CONFINAN	CONJUDGE	CONSCI	CONLEGIS	CONARMY	CONLABOR
Harris 2430	Television news	(X)	The U.S. Supreme Court	(X)	Congress	(X)	(X)
Harris 2434	Television news	(X)	The U.S. Supreme Court	(X)	Congress	The military	Organized labor
Harris 2515	Television news	(X)	(X)	(X)	Congress	The military	Labor unions
GSS 1975	TV	Banks and financial institutions	U.S. Supreme Court	The scientific community	Congress		Organized labor
Harris 7581	TV news	Banks	U.S. Supreme Court	Science	Congress	The military	Labor unions
Harris 7585	Television news	Banks	The U.S. Supreme Court	(X)	Congress	The military	Organized labor

Harris 2521	Television news	Banks	The U.S. Supreme Court	(X)	Congress	The military	Organized labor
Harris 7681	Television news	(X)	The U.S. Supreme Court	(X) of Representa-	The U.S. House of Representatives/U.S. Senate/Congress	The military	(X)
GSS 1976	TV	Banks and financial institutions	U.S. Supreme Court	The scientific community	Congress	Military	Organized labor
Harris 7684	(X)	(X)	U.S. Supreme Court	(X)	Congress	(X)	(X)
Harris 2628	Television news	(X)	(X)	(X)	Congress	The military	Labor unions
Harris 2630	TV news	Banks	(X)	Science	Congress	(X)	(X)
Harris 7690	Television news	Banks	The U.S. Supreme Court	(X)	Congress	The military	Organized labor
GSS 1977	TV	Banks and financial institutions	U.S. Supreme Court	The scientific community	Congress	Military	Organized labor

X Not asked in survey.

business appears first in 25 surveys and immediately precedes religion in 18 surveys. Many similar suborderings frequently occur for other items.

The last way in which the ordering varies is in the number of items asked. The GSS in 1973 and 1974 had 12 items and with the addition of financial confidence in 1975 has had 13 since. NORC 4179 had seven items. The number on Harris varies as follows:

1574	17	2430	8
1702	17	2434	13
2131	17	2515	14
2219	14	7581	37
2236	17	7585	23
2251	20	2521	21
2319	9	7681	17
2343	22	7684	10
2354	32	2628	14
7482	13	2630	12
7487	13	7690	17

As a result, the order of the common Harris/GSS items has depended not only on their relationship to each other but also on the placement and inclusion of a varying number of other confidence items.

The general content of the surveys and the placement of the confidence question was examined. The Harris surveys emphasized political matters, with the exception of Harris 2354, dealing primarily with business matters, and Harris 2131, concerning science. The GSS surveys and NORC amalgam, on the other hand, were eclectic, including a detailed demographic section and attitude and behavioral questions on familial, psychological, social, economic, and politi-

Table 3. Order of Institutions, by Survey

Variable	Harris 1574[a]	Harris 1702[a]	Harris 2131	Harris 2219	Harris 2236	Harris 2251	Harris 2319
CONBUS	1	1	1	1	1	1	1
CONCLERG	2	2	2	(X)	2	2	2
CONEDUC	3	3	3	6	3	3	(X)
CONFED	4	4	4	3	4	4	3
CONLABOR	5	5	5	4	5	5	4
CONPRESS	6	6	6	(X)	6	6	7[b]
CONMEDIC	7	7	7	(X)	7	7	6
CONTV	8	8	8	10	8	8	7[b]
CONFINAN	9	9	9	2	9	9	8
CONJUDGE	11	11	11	(X)	11	11	(X)
CONSCI	12	12	12	(X)	12	12	(X)
CONLEGIS	14	14	14	(X)	14	14	(X)
CONARMY	16	16	16	(X)	16	16	(X)

(See footnotes on p. 131.)

Table 3. Order of Institutions, by Survey—Continued

Variable	GSS 1973	Harris 2343	Harris 2354	NORC 4179	Harris 7482	GSS 1974	Harris 7487
CONBUS	1	1	1	1	1	1	1
CONCLERG	2	2	2	2	(X)	2	2
CONEDUC	3	(3)	(3)	(X)	(X)	3	(3)
CONFED	4	17	4	3	2	4	10
CONLABOR	5	5	5	4	(X)	5	5
CONPRESS	6	6	6	5	(X)	6	6
CONMEDIC	7	8	7	(X)	3	7	7
CONTV	8	(9)	(8)	(X)	(4)	8	(8)
CONFINAN	(X)	(X)	20	(X)	(X)	(X)	(X)
CONJUDGE	9	11	(X)	6	(X)	9	9
CONSCI	10	(X)	9	(X)	(X)	10	(X)
CONLEGIS	11	4/7[c]	10	7	(X)	11	4
CONARMY	12	20	(X)	(X)	(X)	12	11

Variable	Harris 2430	Harris 2434	Harris 2515	GSS 1975	Harris 7581	Harris 7585	Harris 2521
CONBUS	1	1	1	5	1	1	1
CONCLERG	(X)	9	(X)	6	6	2	2
CONEDUC	(X)	(10)	(8)	7	(4)	(3)	(3)
CONFED	7	7	(X)	1	12	11	11
CONLABOR	(X)	11	9	2	15	5	5
CONPRESS	3	3	6	3	7	6	6
CONMEDIC	4	4	(X)	8	37	7	7
CONTV	(5)	(5)	(14)	9	(5)	(9)	(9)
CONFINAN	(X)	(X)	(X)	13	10	8	8
CONJUDGE	6	6	(X)	10	14	10	10
CONSCI	(X)	(X)	(X)	11	3	(X)	(X)
CONLEGIS	2	2	3	12	9	4	4
CONARMY	(X)	12	2	4	2	12	12

Variable	Harris 7681	GSS 1976	Harris 7684	Harris 2628	Harris 2630	Harris 7690[d]	GSS 1977
CONBUS	1	5	(X)	1	1	1	1
CONCLERG	(X)	6	(X)	(X)	(X)	2	2
CONEDUC	(X)	7	(X)	(8)	(X)	(3)	3
CONFED	12	1	1	(X)	7	11	4
CONLABOR	(X)	2	(X)	9	(X)	5	5
CONPRESS	3	3	(X)	6	8	6	6
CONMEDIC	(X)	8	(X)	(X)	3	7	7
CONTV	(4)	9	(X)	(14)	(4)	(9)	8
CONFINAN	(X)	13	(X)	(X)	9	8	13
CONJUDGE	7	10	6	(X)	(X)	10	9
CONSCI	(X)	11	(X)	(X)	2	(X)	10
CONLEGIS	16[e]	12	4	3	5	4	11
CONARMY	14	4	(X)	2	(X)	12	12

Note: Gives absolute position in questionnaire. Parentheses indicate variations in wording.

[a] Instructions read: Rotate asking order and record below.

[b] Items combined.

[c] Item split into parts on House of Representatives (4) and Senate (7).

[d] Instructions read: Use two starting points on list from interview to interview. That is, on first interview start with item 1 on second interview with item 7. On third interview with item 1 again, etc.

[e] Also U.S. Senate is item 2 and the U.S. House of Representatives is item 5.

cal topics. The precise placement of the confidence questions and the content of the questions that immediately preceded them are summarized in table 4. On several of the Harris surveys, the preceding questions had either negative or problem orientations. On Harris 2430, 2434, 2521, and 7487 and NORC 4179, an alienation index offering several pessimistic statements on American society appeared shortly before the confidence question. Harris 2343 asked about the biggest problems facing the country, 1702 asked about the Powell investigation, and 7581 inquired about crises. These types of items might very well lower confidence levels by putting the respondent in a negative frame of mind. The

Table 4. Placement of Confidence Question, by Survey

Survey	Confidence question number	Position of question preceding confidence question	Content of preceding question
Harris 1574	3	1	Complex question with A-M parts and up to 16 subparts to each part; focus on (1) amount of progress in solving listed problems, (2) standard of living, (3) the future of the free enterprise system
Harris 1702	16a	1	Seven-part item on Adam Clayton Powell
		2	15-part item on spending priorities
Harris 2131	8e	1	Four negative and four positive agree/disagree statements on science
		2	"What are the two or three biggest problems you feel science has created as far as you personally are concerned? Any others?"
		3	As above with *benefits* replacing *problems*
Harris 2219	12	1	Four-part question on loan sources
		2	Four-part question on savings accounts
Harris 2236	19	1	Ranking of political philosophy of Nixon, McGovern, Agnew, Shriver, self
		2	10-part item comparing ability of Nixon and McGovern to solve national problems
Harris 2251	1c	1	"In general, over the last ten years, do you feel that America has become a better place to live, a worse place to live, or is just about the way it was ten years ago? What has happened in America over the past ten years to make the country a (better/worse) place to live in? Anything else?"
Harris 2319	1c	1	"In general, over the past ten years, do you feel that America has become a better place to live, a worse place to live, or is just the way it was 10 years ago? (*If better or worse*) What has happened over the past ten years to make the country a (better/worse) place to live in? Anything else?"

(See footnote on p. 134.)

Table 4. Placement of Confidence Question, by Survey—Continued

Survey	Confidence question number	Position of question preceding confidence question	Content of preceding question
Harris 2343	5	1	Four-part question on how local, state, Federal Government affect lives
		2	"What do you feel are the two or three biggest problems facing the country you would like to see something done about? Anything else? What do you think ought to be done about it? Anything else?"
Harris 2354	4a	1	Three-part question about whether business has or should help to solve 20 listed problems
		2	Three-part question rating business contribution to 25 economic goals, comparing confidence to that of 10 years ago
Harris 7482	1	(X)	No prior question, confidence first in survey
Harris 7487	11	1	Five-item alienation index—"Now I want to read you some things some people have told us they have felt from time to time. Do you tend to feel or not (*read list*)? A. The people running the country don't really care what happens to you. B. The rich get richer and the poor get poorer. C. What you think doesn't count very much any more. D. You're left out of things going on around you. E. Most people with power try to take advantage of people like yourself.?
		2	Nine-part question on economic conditions and purchasing plans
Harris 2430	8	1	Five-item alienation index (see Harris 7487)
		2	Presidential choice for 1976
Harris 2434	8	1	Five-item alienation index (see Harris 7487)
		2	Presidential choice for 1976
Harris 2515	4c	1, 2, 3	Energy questions
Harris 7581	2a	1	Three-part questions on access to information, "Do you feel we always have one crisis or another in America, or do you feel there is something deeply wrong in America today?" and "Compared to 10 years ago, do you feel the quality of life in America has improved, grown worse, or stayed the same?"
Harris 7583	1d	1	Three parts on changes over last 10 years in (a) quality of life (see Harris 7581), (b) quality of America as a place to live (see Harris 2319) and (c) quality of leadership (see Harris 2521, part C).

(See footnote on p. 134.)

Table 4. Placement of Confidence Question, by Survey—Continued

Survey	Confidence question number	Position of question preceding confidence question	Content of preceding question
Harris 2521	2d	1	Four-part question, "Compared to 10 years ago, do you feel the leadership inside and outside government in this country has become better, worse, or stayed about the same?", quality of life (see Harris 7581), six-item alienation question (see Harris 7487).
Harris 7681	51a	1	10-part question on federalism
Harris 7684	P6a	1	Complex question with A-L subquestions and up to eight items per letter, focusing on presidential nominating process
Harris 2628	3c	1, 2	Energy questions
Harris 2630	1	1	Party choice in Congressional election
		2, 3	Evaluation of Democratic convention
Harris 7690	2j	1	19-part question on Carter's economic program
NORC 4179	88	1	"Now to something different. I am going to read some of the kinds of things people tell us when we interview them and ask you whether you agree or disagree with them. I'll read them one at a time and you just tell me whether you agree or disagree. A. People like me don't have any say about what the government does. B. I don't think public officials care much what people like me think. C. Generally speaking, those we elect to Congress in Washington lose touch with the people pretty quickly. D. Parties are only interested in people's votes but not in their opinions."
		2	Series of split-ballot policy questions
GSS 1973	56	1	Four-part question about police use of force
		2	Five-part question about citizen use of force
GSS 1974	54	1	List of nine questions on past, present, and future community of residence, preferred type
GSS 1975	44	1	Rate own social class
		2	Rank family income
		3	Change in financial situation over last few years
		4	Satisfaction with financial situation
GSS 1976	1	(X)	No prior question, confidence first in questionnaire
GSS 1977	49	1	Gun ownership
		2	Hunting participation
		3	Ever ticketed or arrested

X Not applicable.

impact of prior questions of the GSS surveys is less certain. GSS 1976 and Harris 7482 had no prior question effect since confidence is the first question, but this fact could have a major impact itself. The questions on use of force in GSS 1973 could have a depressing impact on confidence, but most of the rest appear fairly innocuous. While it is impossible to state with a great certainty whether a context effect might exist, the wide variation in prior questions and the general focus of surveys creates this possibility.

From the preceding discussion of sample population, question wording, format, descriptors, institutional ordering, and context, the Harris and NORC series appear to be a bewildering mixture of similarities and differences. On the similar side, Harris and NORC clearly inquire about the same basic attitude, confidence in the leadership of important institutions. Identical response categories are usually employed, the populations sampled are usually the same, and the descriptors of the institutions are also frequently identical. On the difference side, there are many exceptions to the usual correspondence between sample populations, response categories, and institutional descriptors; multiple variations in wording and format; and many differences in the ordering of institutions and context. Possible consequences of these differences are evaluated after initial inspection of the marginal differences between the Harris and NORC series.

COMPARISON OF DATA

Of the 22 Harris surveys with confidence questions under examination here, raw frequencies were available for 19 studies. The raw data for Harris 1574 and 2131 are lost but published figures exist, and neither raw nor published data survive for Harris 2251. Raw data were available for all NORC surveys (see appendix). The proportion replying "a great deal" appears in table 5.

Table 5. Proportion With a Great Deal of Confidence, by Survey

Variable	Harris 1574	Harris 1702	Harris 2131	Harris 2219	Harris 2236	Harris 2251	Harris 2319
CONBUS	.550	.466	.270	.305	.268	(NA)	.338
CONCLERG	.410	.396	.270	(X)	.294	(NA)	.332
CONEDUC	.610	.555	.370	.310	.334	(NA)	(X)
CONFED	.410	.372	.230	.336	.272	(NA)	(X)
CONLABOR	.220	.196	.140	.103	.153	(NA)	.229
CONPRESS	.290	.265	.180	.165	.184	(NA)	(X)
CONMEDIC	.720	.605	.610	(X)	.482	(NA)	.629
CONTV	.250	.203	.220	.157	.179	(NA)	(X)
CONJUDGE	.500	.395	.230	(X)	.285	(NA)	(X)
CONSCI	.560	.451	.320	(X)	.368	(NA)	(X)
CONLEGIS	.420	.409	.190	(X)	.210	(NA)	(X)
CONARMY	.620	.555	.270	(X)	.361	(NA)	(X)
CONFINAN	.670	.543	.360	.591	.391	(NA)	.499

(See footnotes on p. 136.)

Table 5. Proportion With a Great Deal of Confidence, by Survey--Continued

Variable	GSS 1973	Harris 2343	Harris 2354	NORC 4179	Harris 7482	GSS 1974	Harris 7487
CONBUS	.293	.298	.276	.218	.241	.314	.217
CONCLERG	.348	.356	.289	.321	(X)	.443	.318
CONEDUC	.370	.442	.455	(X)	(X)	.491	.391
CONFED	.293	.194	.134	.142	.117	.136	.283
CONLABOR	.155	.198	.162	.187	(X)	.182	.174
CONPRESS	.231	.303	.278	.251	(X)	.259	.248
CONMEDIC	.541	.576	.599	(X)	.526	.604	.493
CONTV	.186	.403	.366	(X)	.342	.234	(X)
CONJUDGE	.315	.333	(X)	.341	(X)	.332	.401
CONSCI	.369	(X)	.455	(X)	(X)	.450	(X)
CONLEGIS	.235	.297	.171	.227	(X)	.171	.178
CONARMY	.317	.405	(X)	(X)	(X)	.396	.339
CONFINAN	(X)	(X)	.412	(X)	(X)	(X)	(X)

Variable	Harris 2430	Harris 2434	Harris 2515	GSS 1975	Harris 7581	Harris 7585	Harris 2521
CONBUS	.152	.159	.181	.193	.197	.197	.163
CONCLERG	(X)	.320	(X)	.244	.322	.355	.237
CONEDUC	(X)	.393	.362	.309	.361	.366	.279
CONFED	.200	.177	(X)	.133	.131	.160	.108
CONLABOR	(X)	.185	.163	.101	.135	.180	.099
CONPRESS	.309	.256	(X)	.239	.259	.275	.201
CONMEDIC	.497	.485	(X)	.505	.428	.537	.420
CONTV	.362	.323	.336	.178	(X)	.366	.279
CONJUDGE	.348	.350	(X)	.308	.287	.275	.219
CONSCI	(X)	(X)	(X)	.377	.479	(X)	(X)
CONLEGIS	.162	.164	.124	.133	.136	.121	.088
CONARMY	(X)	.307	.267	.352	.245	.303	.225
CONFINAN	(X)	(X)	(X)	.319	.415	.423	.335

Variable	Harris 7681	GSS 1976	Harris 7684	Harris 2628	Harris 2630	Harris 7690	GSS 1977
CONBUS	.215	.220	(X)	.205	.199	.204	.272
CONCLERG	(X)	.307	(X)	(X)	(X)	.293	.400
CONEDUC	(X)	.375	(X)	.317	(X)	.370	.406
CONFED	.165	.135	.223	(X)	.145	.233	.279
CONLABOR	(X)	.116	(X)	.106	(X)	.145	.148
CONPRESS	.213	.285	(X)	.250	.247	.178	.251
CONMEDIC	(X)	.541	(X)	(X)	.501	.425	.515
CONTV	.283	.187	(X)	.326	.345	.276	.174
CONJUDGE	.316	.354	.379	(X)	(X)	.286	.357
CONSCI	(X)	.429	(X)	(X)	.444	(X)	.410
CONLEGIS	.179	.137	.167	.095	.127	.165	.191
CONARMY	.362	.392	(X)	.304	(X)	.276	.363
CONFINAN	(X)	.395	(X)	(X)	.360	.400	.419

NA Not available.
X Not asked in survey.

At six points Harris and NORC surveys were conducted at sufficiently close times (about 1 month apart) to permit direct survey-to-survey comparisons (these comparisons do not eliminate the possibility of real across-time changes between surveys but at least tend to minimize this factor). The comparisons were between (1) Harris 2319 and GSS 1973 (GSS adapted to approximate Harris universe of electoral participators); (2) Harris 2354, NORC 2179, Harris 7482, and GSS 1974; (3) Harris 2525, GSS 1975, and Harris 7581; (4) Harris 7487, 2430, and 2434; (5) Harris 2521, 7681, and GSS 1975; and (6) Harris 7684, 2628, and 2630. On these six comparisons between 18 surveys, a total of 93 comparisons between items were possible. There were 45 comparisons between NORC and Harris surveys, 41 between different Harris surveys, and 7 between NORC surveys. Table 6 shows the difference in proportions between these 93 pairs of marginals and tests for their statistical significance. With a few exceptions, all of these pairs of comparisons were between identical or very similar descriptors and similar sample populations. The chief exceptions are that Harris 2515, 7581, and 2628 used *labor unions* while Harris 7690 and GSS employed *organized labor*. Also, Harris 7684 and 2630 were samples of electoral participators and Harris 2628 sampled adults in general. The evidence indicates little difference attributable to these variations.

Between NORC and Harris .511 of the differences were significant, between Harris and Harris .366 were significant, and between the few NORC-to-NORC comparisons .429 were significant. The average absolute differences in proportions were Harris-NORC = .048, Harris-Harris = .037, and NORC-NORC = .043 (or interhouse = .048, intrahouse = .038). Both in terms of the proportion of significant differences and the magnitude of the average absolute differences there is considerable variation between surveys. By far the largest inter- and intrahouse differences occur respectively between Harris 2521 and GSS 1976 and Harris 2521 and Harris 7681. Eight of the ten items differ significantly between Harris 2521 and GSS 1976 (average difference = .079) and five of the six items differ between Harris 2521 and Harris 7681 (average difference = .074). By contrast, GSS 1976 and Harris 7681 had only two out of six items significantly differing (average difference = .036). Likewise across the other five points of comparison only .286 of Harris-Harris differences vary significantly and only .448 of Harris-GSS differences are significant. This of course suggests that Harris 2521 is the source of atypically large variations between surveys.

Even without these especially large variations, it appears that both within houses and across houses the confidence items often vary significantly within a relatively short time span. How much results from real fluctuations in confidence ratings and how much from artificial differences in context, wording, and so forth is difficult to ascertain. If we compare the mean interhouse difference (.048) with the mean intrahouse difference (.038), on average we find that items differ by a percentage point more between Harris and NORC surveys than between surveys conducted by the same house. While this comparison is hardly experimentally rigorous, it probably accurately reflects the fact that a combination of form differences in the items and more basic differences in house

Table 6. Differences Between Contiguous Surveys

Variable	Harris 2319	GSS 1973	Harris 2319-GSS 1973		Harris 2354[a]	NORC 4179
			d	*p*		
CONBUS	.338	.325	.013	.650	.276	.218
CONCLERG	.332	.331	.001	.971	.289	.321
CONEDUC	(X)	(X)	(X)	(X)	(X)	(X)
CONFED	(X)	(X)	(X)	(X)	.134	.142
CONLABOR	.229	.142	.087	<.001	.162	.187
CONPRESS	(X)	(X)	(X)	(X)	.278	.251
CONMEDIC	.629	.553	.076	.010	(X)	(X)
CONTV	(X)	(X)	(X)	(X)	(X)	(X)
CONJUDGE	(X)	(X)	(X)	(X)	(X)	.341
CONSCI	(X)	(X)	(X)	(X)	(X)	(X)
CONLEGIS	(X)	(X)	(X)	(X)	.171	.227
CONARMY	(X)	(X)	(X)	(X)	(X)	(X)
CONFINAN	(X)	(X)	(X)	(X)	(X)	(X)

Variable	Harris 7482	GSS 1974[a]	Harris 7482-NORC 4179		GSS 1974-NORC 4179	
			d	*p*	*d*	*p*
CONBUS	.241	.314	.023	.293	.096	<.001
CONCLERG	(X)	.443	(X)	(X)	.122	<.001
CONEDUC	(X)	(X)	(X)	(X)	(X)	(X)
CONFED	.117	.136	−.025	.148	−.006	.738
CONLABOR	(X)	.182	(X)	(X)	−.005	.800
CONPRESS	(X)	.259	(X)	(X)	.008	.724
CONMEDIC	(X)	(X)	(X)	(X)	(X)	(X)
CONTV	(X)	(X)	(X)	(X)	(X)	(X)
CONJUDGE	(X)	.332	(X)	(X)	−.009	.715
CONSCI	(X)	(X)	(X)	(X)	(X)	(X)
CONLEGIS	(X)	.171	(X)	(X)	−.056	.007
CONARMY	(X)	(X)	(X)	(X)	(X)	(X)
CONFINAN	(X)	(X)	(X)	(X)	(X)	(X)

Variable	GSS 1974-Harris 7482		NORC 4179-Harris 2354		Harris 7482-Harris 2354	
	d	*p*	*d*	*p*	*d*	*p*
CONBUS	.073	.002	−.058	.009	−.035	.120
CONCLERG	(X)	(X)	.032	.177	(X)	(X)
CONEDUC	(X)	(X)	(X)	(X)	(X)	(X)
CONFED	.019	.271	.008	.660	−.017	.675
CONLABOR	(X)	(X)	.025	.202	(X)	(X)
CONPRESS	(X)	(X)	−.027	.237	(X)	(X)
CONMEDIC	(X)	(X)	(X)	(X)	(X)	(X)
CONTV	(X)	(X)	(X)	(X)	(X)	(X)
CONJUDGE	(X)	(X)	(X)	(X)	(X)	(X)
CONSCI	(X)	(X)	(X)	(X)	(X)	(X)
CONLEGIS	(X)	(X)	.056	.007	(X)	(X)
CONARMY	(X)	(X)	(X)	(X)	(X)	(X)
CONFINAN	(X)	(X)	(X)	(X)	(X)	(X)

(See footnotes on p. 141.)

Table 6. Differences Between Contiguous Surveys—Continued

Variable	Harris 2515	GSS 1975	Harris 7581	Harris 2515-GSS 1975	
				d	*p*
CONBUS	.181	.193	.197	−.011	.541
CONCLERG	(X)	.244	.322	(X)	(X)
CONEDUC	(X)	(X)	(X)	(X)	(X)
CONFED	(X)	.133	.131	(X)	(X)
CONLABOR	.163	.101	.135	.062	.001
CONPRESS	(X)	.239	.259	(X)	(X)
CONMEDIC	(X)	.505	.428	(X)	(X)
CONTV	(X)	(X)	(X)	(X)	(X)
CONJUDGE	(X)	.308	.287	(X)	(X)
CONSCI	(X)	.377	.479	(X)	(X)
CONLEGIS	.124	.133	.136	−.009	.593
CONARMY	.267	.352	.245	−.085	.001
CONFINAN	(X)	.319	.415	(X)	(X)

Variable	Harris 7581-GSS 1975		Harris 2515-Harris 7581		Harris 7487	Harris 2430
	d	*p*	*d*	*p*		
CONBUS	.004	.838	−.016	.595	.217	.159
CONCLERG	.078	.001	(X)	(X)	.318	(X)
CONEDUC	(X)	(X)	(X)	(X)	(X)	(X)
CONFED	−.002	.904	(X)	(X)	.283	.200
CONLABOR	.034	.036	.028	.101	.174	(X)
CONPRESS	.020	.631	(X)	(X)	.248	.309
CONMEDIC	−.077	.003	(X)	(X)	.493	.497
CONTV	(X)	(X)	(X)	(X)	(X)	.362
CONJUDGE	−.021	.627	(X)	(X)	.401	.348
CONSCI	.102	<.001	(X)	(X)	(X)	(X)
CONLEGIS	.003	.858	−.012	.530	.178	.162
CONARMY	−.107	<.001	−.022	.299	.339	(X)
CONFINAN	.096	<.001	(X)	(X)	(X)	(X)

Variable	Harris 2434	Harris 2430-Harris 7487		Harris 2434-Harris 7487	
		d	*p*	*d*	*p*
CONBUS	.152	−.058	.010	−.065	.004
CONCLERG	.320	(X)	(X)	.002	.931
CONEDUC	(X)	(X)	(X)	(X)	(X)
CONFED	.177	−.083	<.001	−.106	<.001
CONLABOR	.185	(X)	(X)	.011	.584
CONPRESS	.256	.061	.044	.008	.720
CONMEDIC	.485	.004	.902	−.008	.754
CONTV	.323	(X)	(X)	(X)	(X)
CONJUDGE	.350	−.053	.100	−.051	.037
CONSCI	(X)	(X)	(X)	(X)	(X)
CONLEGIS	.164	−.016	.534	−.014	.524
CONARMY	.307	(X)	(X)	.032	.179
CONFINAN	(X)	(X)	(X)	(X)	(X)

(See footnotes on p. 141.)

Table 6. Differences Between Contiguous Surveys—Continued

Variable	Harris 2434-Harris 2430		Harris 2521	Harris 7681	GSS 1976
	d	*p*			
CONBUS	—.007	.772	.163	.215	.220
CONCLERG	(X)	(X)	.237	(X)	.307
CONEDUC	(X)	(X)	(X)	(X)	(X)
CONFED	—.023	.606	.108	.165	.135
CONLABOR	(X)	(X)	.099	(X)	.116
CONPRESS	—.053	.081	.201	.213	.285
CONMEDIC	—.012	.724	.420	(X)	.541
CONTV	—.039	.226	(X)	(X)	(X)
CONJUDGE	.002	.949	.219	.316	.354
CONSCI	(X)	(X)	(X)	(X)	(X)
CONLEGIS	—.002	.934	.088	.179	.137
CONARMY	(X)	(X)	.225	.362	.392
CONFINAN	(X)	(X)	.335	(X)	.395

Variable	Harris 7681-Harris 2521		GSS 1976-Harris 2521		GSS 1976-Harris 7681	
	d	*p*	*d*	*p*	*d*	*p*
CONBUS	.052	.010	.057	.005	.005	.809
CONCLERG	(X)	(X)	.070	.003	(X)	(X)
CONEDUC	(X)	(X)	(X)	(X)	(X)	(X)
CONFED	.057	.002	.027	.106	—.030	.099
CONLABOR	(X)	(X)	.017	.289	(X)	(X)
CONPRESS	.012	.573	.084	<.001	.072	.002
CONMEDIC	(X)	(X)	.121	<.001	(X)	(X)
CONTV	(X)	(X)	(X)	(X)	(X)	(X)
CONJUDGE	.097	<.001	.135	<.001	.038	.114
CONSCI	(X)	(X)	(X)	(X)	(X)	(X)
CONLEGIS	.091	<.001	.049	.003	—.042	.024
CONARMY	.137	<.001	.167	<.001	.030	.227
CONFINAN	(X)	(X)	.060	.015	(X)	(X)

Variable	Harris 7684	Harris 2628	Harris 2630	Harris 2628-Harris 7684	
				d	*p*
CONBUS	(X)	.205	.199	(X)	(X)
CONCLERG	(X)	(X)	(X)	(X)	(X)
CONEDUC	(X)	(X)	(X)	(X)	(X)
CONFED	.223	(X)	.145	(X)	(X)
CONLABOR	(X)	(X)	(X)	(X)	(X)
CONPRESS	(X)	.250	.247	(X)	(X)
CONMEDIC	(X)	(X)	(X)	(X)	(X)
CONTV	(X)	.326	.345	(X)	(X)
CONJUDGE	(X)	(X)	(X)	(X)	(X)
CONSCI	(X)	(X)	(X)	(X)	(X)
CONLEGIS	.167	.095	.127	—.072	<.001
CONARMY	(X)	(X)	(X)	(X)	(X)
CONFINAN	(X)	(X)	(X)	(X)	(X)

(See footnotes on p. 141.)

Table 6. Differences Between Contiguous Surveys—Continued

Variable	Harris 2630-Harris 7684		Harris 2630-Harris 2628	
	d	*p*	*d*	*p*
CONBUS	(X)	(X)	−.006	.759
CONCLERG	(X)	(X)	(X)	(X)
CONEDUC	(X)	(X)	(X)	(X)
CONFED	−.078	<.001	(X)	(X)
CONLABOR	(X)	(X)	(X)	(X)
CONPRESS	(X)	(X)	−.003	.882
CONMEDIC	(X)	(X)	(X)	(X)
CONTV	(X)	(X)	.019	.582
CONJUDGE	(X)	(X)	(X)	(X)
CONSCI	(X)	(X)	(X)	(X)
CONLEGIS	−.040	.028	.032	.036
CONARMY	(X)	(X)	(X)	(X)
CONFINAN	(X)	(X)	(X)	(X)

Note: To adjust for multistage sampling, standard deviations were multiplied by 1.414, and probabilities were calculated from these modified figures.

X Not applicable.

d Difference.

p Probability.

[a]Harris 2354 and GSS 1974 are not contiguous in time, and no comparison is made.

procedures (e.g., sample frame, multistage procedures, interviewer training) create an added measure of variation between Harris and NORC on these confidence items.

To examine further the similarities and differences between the Harris and NORC data, a comparison of Harris and NORC trends from 1972 to 1977 was made. Taking a conservative approach, Harris surveys that sampled electoral participators (2236, 7684, and 2630) or employed persons (Harris 2319) or that used institutional descriptors judged to be major variants (large business corporations, full service banks, newspapers, Federal Government in Harris 2219; labor unions in Harris 2219, 2515, 7581, and 2628; the U.S. House of Representatives and the U.S. Senate in Harris 2343; and science in Harris 7581 and 2630) were dropped from the initial time series comparisons between Harris and NORC (tables 7 and 8). This procedure permitted comparison of Harris and GSS trends on 10 institutions—Congress, the U.S. Supreme Court, the executive branch, organized religion, medicine, the press, organized labor, the military, major companies, and banks.

The trend comparisons were also hampered by the differences in the time points covered. The two series often started and/or ended several months apart and of course usually covered different times within a span of years. The possible impact on trends of these differences in coverage is considered in particular cases.

To evaluate the trends, first no-change, or constant, models were fitted to the

Table 7. Trends in Proportion With a Great Deal of Confidence, by Survey and House: 1972-1977

Variable and house	Model	χ^2	df	Probability	Decision
CONJUDGE					
Harris	$p = c$	159.3	8	<.001	R
	$p = a + bx$	102.3	7	<.001	R
	LR	57.3	1	<.001	S
NORC	$p = c$	13.7	5	.018*	A
CONARMY					
Harris	$p = c$	197.2	9	<.001	R
	$p = a + bx$	155.9	8	<.001	R
	LR	41.3	1	<.001	S
NORC	$p = c$	27.4	4	<.001	R
	$p = a + bx$	21.8	3	<.001	R
	LR	5.6	1	.017*	NS
CONLEGIS					
Harris	$p = c$	154.5	10	<.001	R
	$p = a + bx$	126.7	9	<.001	R
	LR	27.8	1	<.001	S
NORC	$p = c$	96.5	5	<.001	R
	$p = a + bx$	72.7	4	<.001	R
	LR	23.8	1	<.001	S
CONMEDIC					
Harris	$p = c$	219.3	9	<.001	R
	$p = a + bx$	83.0	8	<.001	R
	LR	136.2	1	<.001	S
NORC	$p = c$	36.4	4	<.001	R
	$p = a + bx$	28.2	3	<.001	R
	LR	8.2	1	.004	S
CONPRESS					
Harris	$p = c$	64.2	10	<.001	R
	$p = a + bx$	22.6	9	<.001*	A
	LR	41.6	1	<.001	S
NORC	$p = c$	13.5	5	.019*	A
CONCLERG					
Harris	$p = c$	79.3	8	<.001	R
	$p = a + bx$	63.4	7	<.001	R
	LR	16.0	1	<.001	S
NORC	$p = c$	171.1	5	<.001	R
	$p = a + bx$	172.5	4	<.001	R
	LR	1.3	1	1.000	NS
CONFED					
Harris	$p = c$	267.6	11	<.001	R
	$p = a + bx$	283.7	10	<.001	R
	LR	−16.1	1	1.000	NS
NORC	$p = c$	256.9	5	<.001	R
	$p = a + bx$	272.9	4	<.001	R
	LR	−15.9	1	1.000	NS
CONLABOR					
Harris	$p = c$	91.3	6	<.001	R
	$p = a + bx$	53.4	5	<.001	R
	LR	37.8	1	<.001	S
NORC	$p = c$	75.4	5	<.001	R
	$p = a + bx$	61.0	4	<.001	R
	LR	14.5	1	<.001	S

(See footnotes on p. 143.)

Table 7. Trends in Proportion With a Great Deal of Confidence, by Survey and House: 1972-1977—Continued

Variable and house	Model	χ^2	df	Probability	Decision
CONBUS					
Harris	$p = c$	181.0	12	<.001	R
	$p = a + bx$	142.7	11	<.001	R
	LR	38.3	1	<.001	S
NORC	$p = c$	92.5	5	<.001	R
	$p = a + bx$	88.3	4	<.001	R
	LR	4.2	1	.038*	NS
CONFINAN[a]					
Harris	$p = c$	31.4	3	<.001	R
	$p = a + bx$	31.4	2	<.001	R
	LR	0	1	1.000	NS
NORC	$p = c$	36.1	2	<.001	R
	$p = a + bx$	2.9	1	<.087	A

Note:

p = proportion
c = constant
df = degrees of freedom
R = reject
A = accept
S = significant at .05
NS = not significant at .05
LR = linear reduction

[a]1975-1977 only.
*Not significant at .05 when adjusted for multistage sampling.

separate GSS and Harris series. If the model proved inadequate to explain the series, a linear-change model was fitted to the marginals. The results of these tests are given in table 7.

Taking the Harris military points as an example, the constant hypothesis is rejected because there is a significant amount of variation (χ^2) unexplained by a constant fit. The linear hypothesis is likewise rejected because a significant amount of χ^2 again remains unexplained by the best linear fit. However, the linear reduction (the amount of χ^2 unexplained by the constant model minus the amount unexplained by the linear model) is significant, which indicates that although a simple linear model does not adequately fit the data, there is a significant linear component in the more complex trend. In other words, the figures bounce too much to be linear, but the bouncing has a direction to it. The NORC military series shows another possible outcome. Here neither the constant nor the linear model fits the data, and the linear model is not a significant improvement over the constant model. This represents a nonlinear trend. In brief, for each of the series there are four possible evaluations of the trends:

1. Constant
2. Linear

Table 8. Comparison of Trends for Harris and GSS: 1972-1977

Variable and house	$p = c$	$p = a + bx$	Linear reduction	Slope	$b_h - b_g$ probability	Pooled p	$pp_h - pp_g$ probability
CONLEGIS							
Harris	R	R	S	−.0192	NS	.135	< .001
NORC	R	R	S	−.0157		.176	
CONJUDGE							
Harris	R	R	S	−.0320	(X)	.306	.003
NORC	A	(X)	(X)	(X)		.334	
CONFED							
Harris	R	R	NS	(X)	(X)	.161	.187
NORC	R	R	NS	(X)		.170	
CONCLERG							
Harris	R	R	S	−.0160	(X)	.309	.002
NORC	R	R	NS	(X)		.338	
CONMEDIC							
Harris	R	R	S	−.0476	NS	.499	< .001
NORC	R	R	S	−.0117		.542	
CONPRESS							
Harris	R	A	(X)	−.0296	(X)	.245	< .539
NORC	A	(X)	(X)	(X)		.252	
CONLABOR							
Harris	R	R	S	−.0198	NS	.158	.024
NORC	R	R	S	−.0113		.142	
CONARMY							
Harris	R	R	S	−.0256	(X)	.298	< .001
NORC	R	R	NS	(X)		.363	
CONBUS							
Harris	R	R	S	−.0184	(X)	.205	< .001
NORC	R	R	NS	(X)		.247	
CONFINAN[a]							
Harris	R	R	NS	(X)	(X)	.392	.233
NORC	R	A	(X)	.0504		.376	

Note: See table 7 for nomenclature.
X Not applicable.
[a]1975-1977 only.

3. Linear component
4. Nonlinear

(For further details on the methods used here, see Taylor, 1976.)

Comparisons between the Harris and NORC series were made in several ways. First, they were compared on the type of trend that fits each series. Second, they were compared on their pooled proportion. Last, for those series that tested as linear or linear component, their slopes were calculated. When both Harris and NORC showed linearity, a test was made to see if their slopes differed.

In general, the two series showed a fairly wide degree of divergence. On NORC, two items tested as constant, one as linear, three as linear components, and four as nonlinear. On Harris, none were constant, one was linear, seven were

linear components, and two were nonlinear. As we see below, in only 4 of 10 cases did the data model in a similar fashion for Harris and NORC:

NORC items	Harris items				Total
	Constant	Linear	Linear component	Nonlinear	
Constant	0	1	1	0	2
Linear	0	0	0	1	1
Linear component	0	0	3	0	3
Nonlinear	0	0	3	1	4
Total	0	1	7	2	10

Similarly, when the pooled proportions were compared, in only 3 out of 10 comparisons were the differences statistically insignificant (table 8). In looking at the slopes on the three series that showed linearity (Congress, medicine, and organized labor), no significant differences were found in the slopes, although this resulted as much from the weakness of the linear fits (and thus large standard deviations) as from the proximity of the slopes.[2] Using the Harris series with variants included changes the trend fit for several Harris items but results in about the same degree of matching with the NORC series. Three items out of 11 test out to similar models; 2 of 11 items do not significantly differ in their pooled proportions, and the 2 items that show linearity on both series (Congress and medicine) do not significantly differ in their slopes.

It is, of course, highly probable that the inclusion of the variant populations and descriptors surveys add differences attributable to these particular variations. Since the demonstrable differences in these surveys did not appear to be clearly excessive, it was decided to test how the Harris variant series with its additional data points and slightly different time overlap compared to the Harris standard series in matching with the NORC series. Evidence on the possible usability of the electoral participators surveys comes from two sources. The 1973 GSS was adjusted to eliminate nonparticipators, and none of the modified marginals differed from the unmodified GSS sample by a significant degree (maximum change was only 1.5 percent). Also, Harris 2628 (an adult sample) did not significantly differ from Harris 7684 and 2630 in four out of five comparisons (table 6). On the institutional descriptors there is less evidence to judge how much the marginals might be affected by the variants. With the exception of full service banking, none produced marginals that were incontestably at odds with the standard versions (since this variant appeared in a 1972 survey and Harris and GSS had parallel series only from 1975 to 1977 on the banking items, this variant was automatically dropped from further consideration). The Congress/Senate version on Harris 2343 also appeared to be suspici-

[2] The r^2 between the NORC and Harris series and time were relatively modest: Congress = .292 (NORC), .165 (Harris); medicine = .223 (NORC), .612 (Harris); and organized labor = .241 (NORC), .412 (Harris).

ously high, but when U.S. Senate, the U.S. House of Representatives, and Congress were all asked on Harris 7681, their marginals were close (.193, .196, and .179, respectively). Because of this similarity the Harris 2343 marginals cannot be clearly dismissed as a result of the variant descriptor. On the executive branch, business, and science, the marginals from the variant descriptors are plausible given the GSS and Harris trends, but it is really impossible to know how much they may vary from the standard version. On labor there are actually two series, organized labor with seven Harris points and labor unions with four points. The standard version had a pooled average proportion of .158, significantly higher than the .119 on the variant wording (as the Procter & Gamble data also suggest). Inspection of the time series reveals, however, that it is not possible to rule out that the differences come from the temporal occurrence of the surveys.

Next, each of the series were considered on a case-by-case basis. The trends are depicted in figures 1-11 for NORC series and the standard Harris series with the variant Harris points connected to the standard Harris points. Figure 1 shows that both series find a decline in confidence in the Congress with a partial recovery in 1977. Both the Harris series and the NORC series show a declining linear component, and their slopes do not differ significantly. NORC and Harris differ in that pooled confidence is significantly higher for NORC than for Harris (.041 above the standard, .029 above the variant pooled proportions).

On science (figure 2), the Harris variant series (no standard series exists) shows

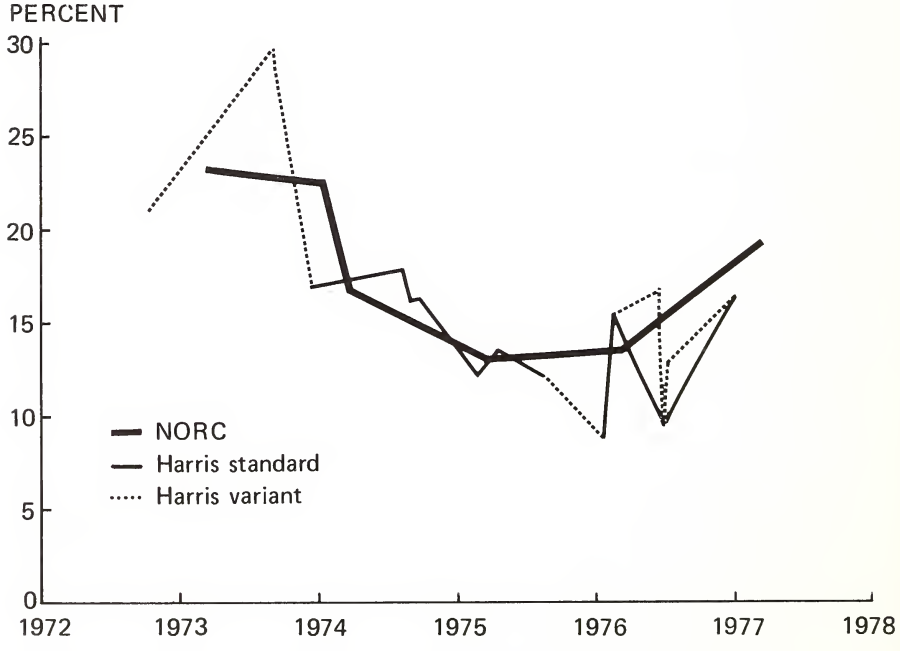


Figure 1. Great deal of confidence in Congress.

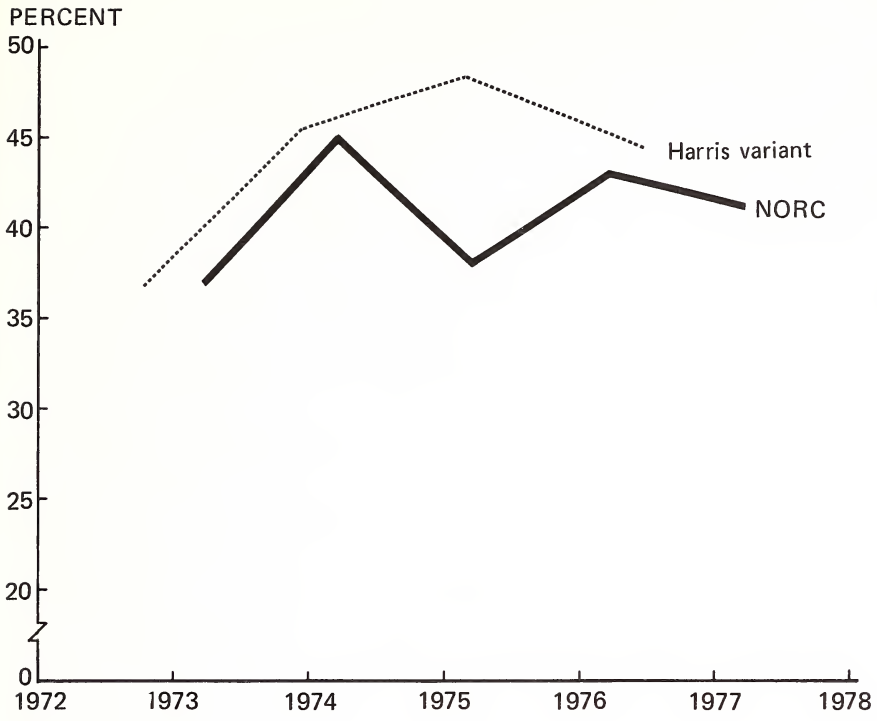


Figure 2. Great deal of confidence in the scientific community.

a linear component increase in confidence while the NORC series is nonlinear. Harris also has a higher pooled confidence than NORC (Harris-NORC = .029).

On the U.S. Supreme Court (figure 3), the Harris standard shows a linear component decline, the Harris variant is nonlinear, and NORC is constant. NORC averages slightly more confidence than the Harris series (standard = .024; variant = .028). NORC also differs in that it shows considerably less variability than the Harris series.

For the executive branch (figure 4), the NORC series shows a nonlinear U-shaped trend while Harris has a W-shaped trend (nonlinear on the standard, linear component decline on the variant). The middle peak on the Harris W results from a survey taken after Richard Nixon's resignation and before his pardon by Gerald Ford, a point of sharp and very short-lived confidence in the Presidency. Immediately after the pardon, confidence began to plummet back to preresignation levels (Smith and Taylor, 1980). With this episodic effect discounted, the Harris and GSS series follow a similar nonlinear U-shaped pattern. On their pooled proportions, NORC and the Harris standard series do not significantly vary, while the Harris variant series shows more confidence than NORC (Harris-NORC = .014). This difference results from two Harris surveys before the Watergate disclosures.

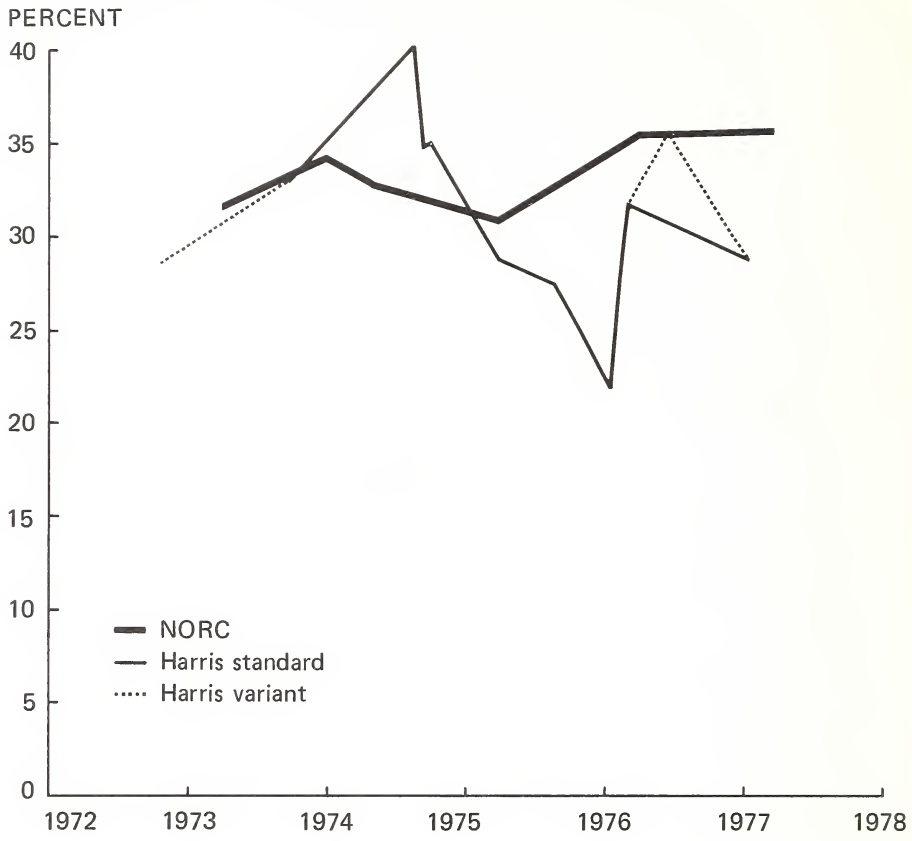


Figure 3. Great deal of confidence in U.S. Supreme Court.

Figure 5, on organized religion, shows a nonlinear trend with wide annual fluctuations for NORC, the Harris standard has a weak linear component decline, and the Harris variant is nonlinear. Harris also finds slightly less overall confidence than NORC does.

On medicine (figure 6), Harris and NORC both manifest linear component declines with no significant differences between their slopes. Each series shows a high degree of variation from this trend, however. NORC records a higher pooled level of confidence than Harris (standard = –.043; variant = –.045).

NORC is constant on the press (figure 7), the Harris standard shows a linear decline, and the Harris variant is nonlinear. NORC and Harris standards do not differ significantly in their pooled proportions, but the Harris variant reports less confidence (Harris-NORC = .026).

On organized labor (figure 8), NORC and Harris standards both show linear component declines with slopes that do not significantly differ, while the Harris variant is nonlinear. NORC shows slightly more pooled confidence than the

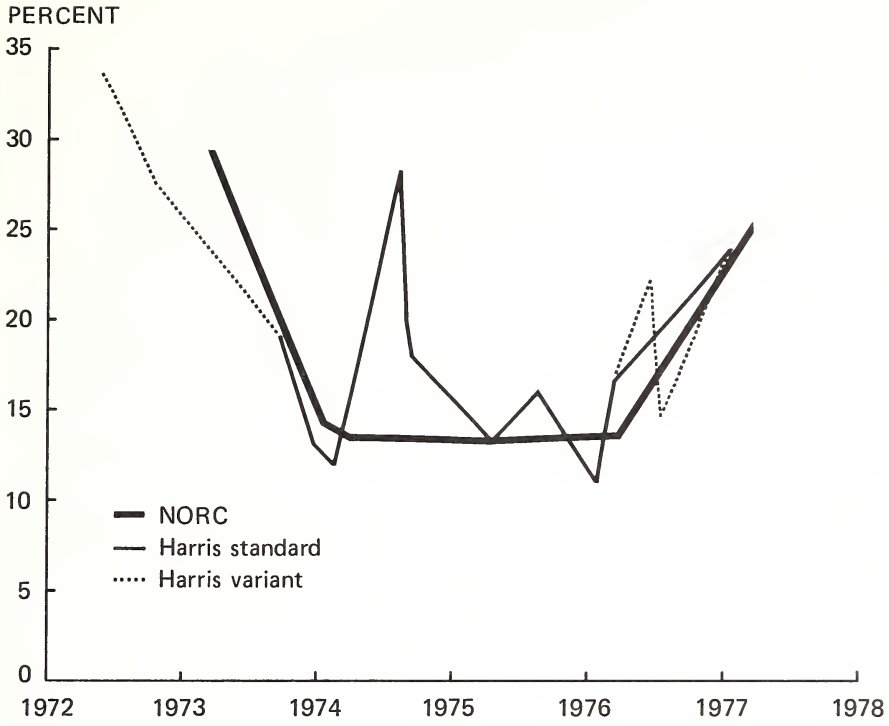


Figure 4. Great deal of confidence in the executive branch.

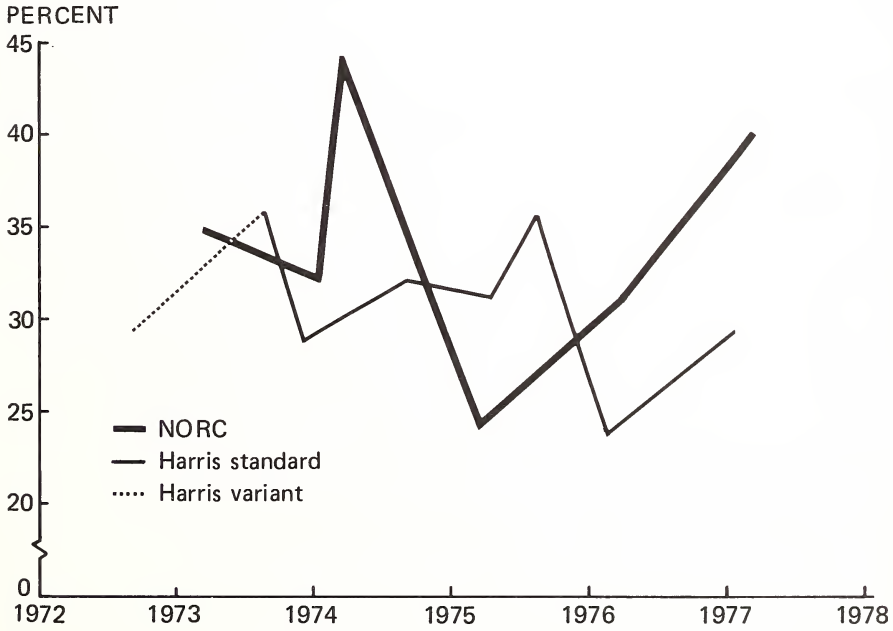


Figure 5. Great deal of confidence in organized religion.

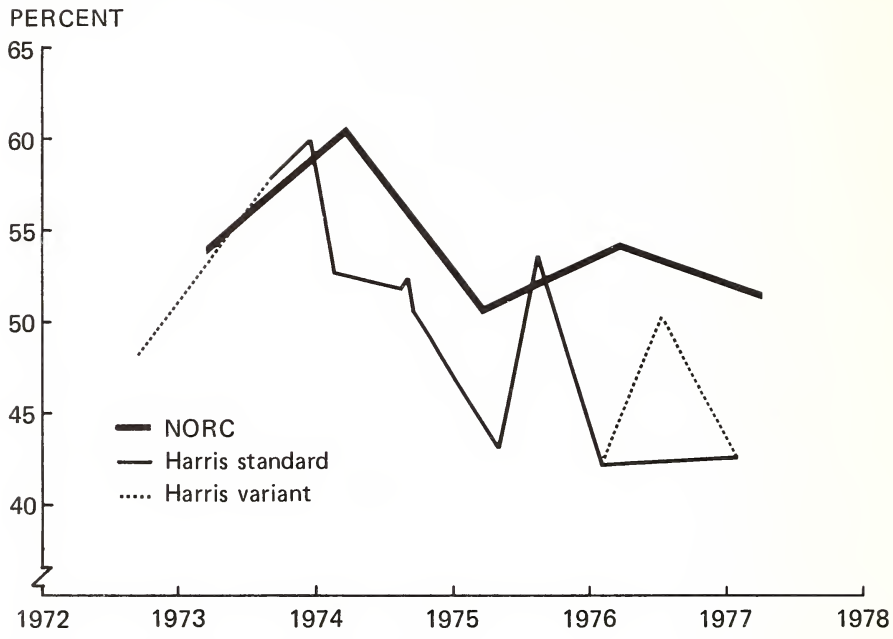


Figure 6. Great deal of confidence in medicine.

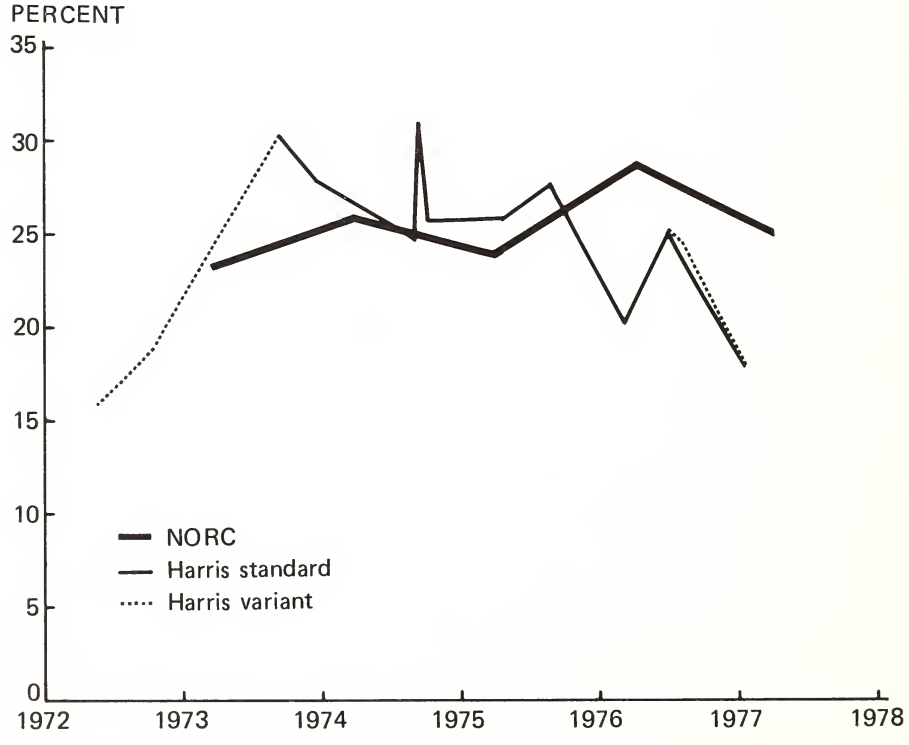


Figure 7. Great deal of confidence in the press.

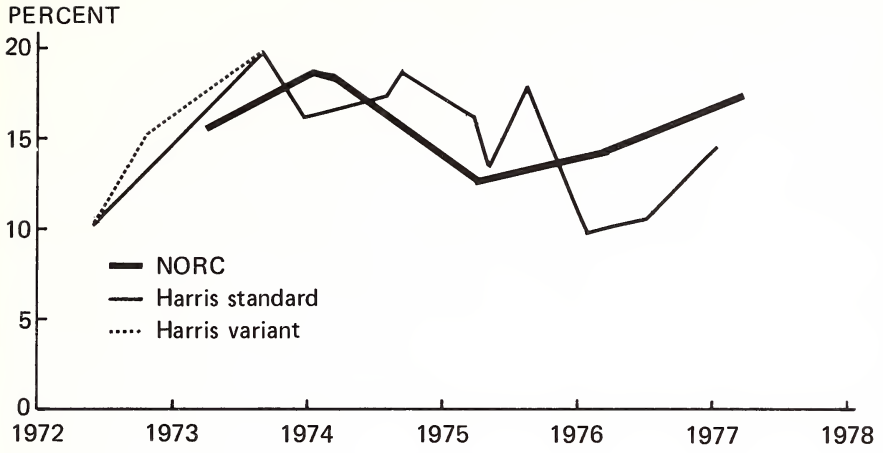


Figure 8. Great deal of confidence in organized labor.

Harris standard (Harris-NORC = .016), but NORC and Harris variants do not differ significantly.

On the military (figure 9), NORC has a nonlinear trend (but with a linear component increase of .0092 per annum of borderline significance), while the Harris series show linear component declines. Both Harris series record much less confidence than NORC does (standard = —.065; variant = —.060).

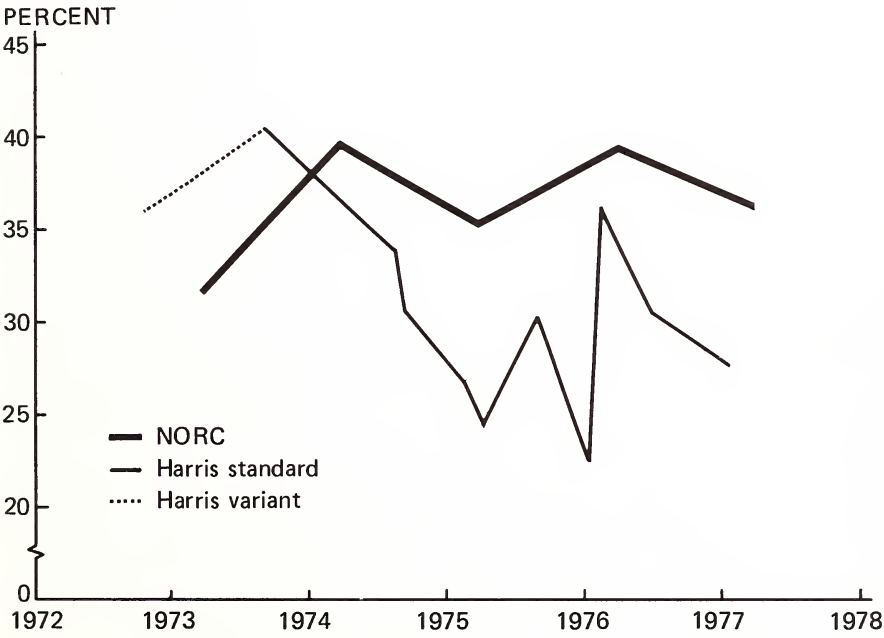


Figure 9. Great deal of confidence in the military.

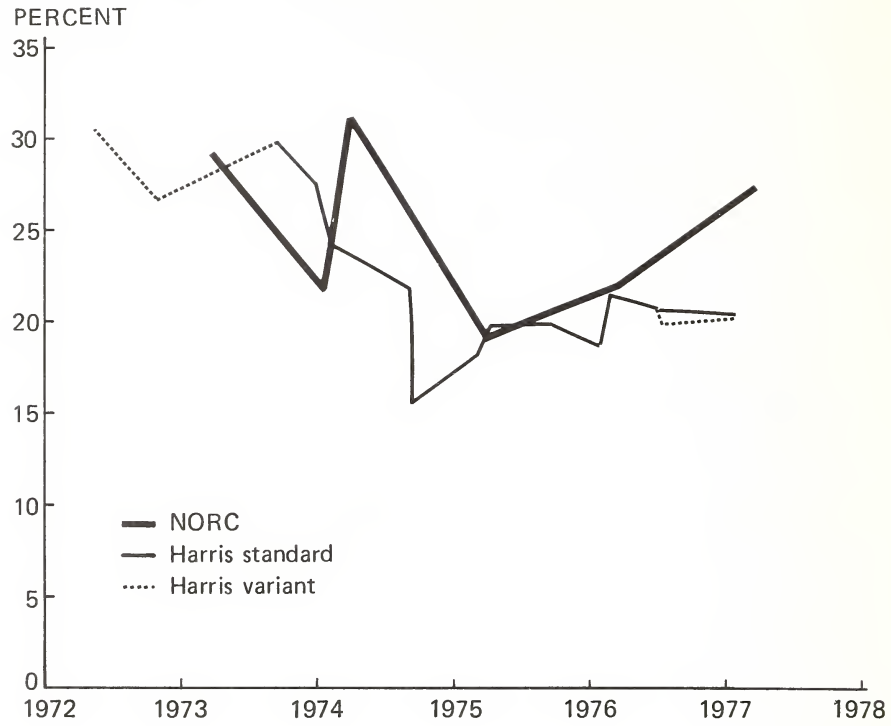


Figure 10. Great deal of confidence in major companies.

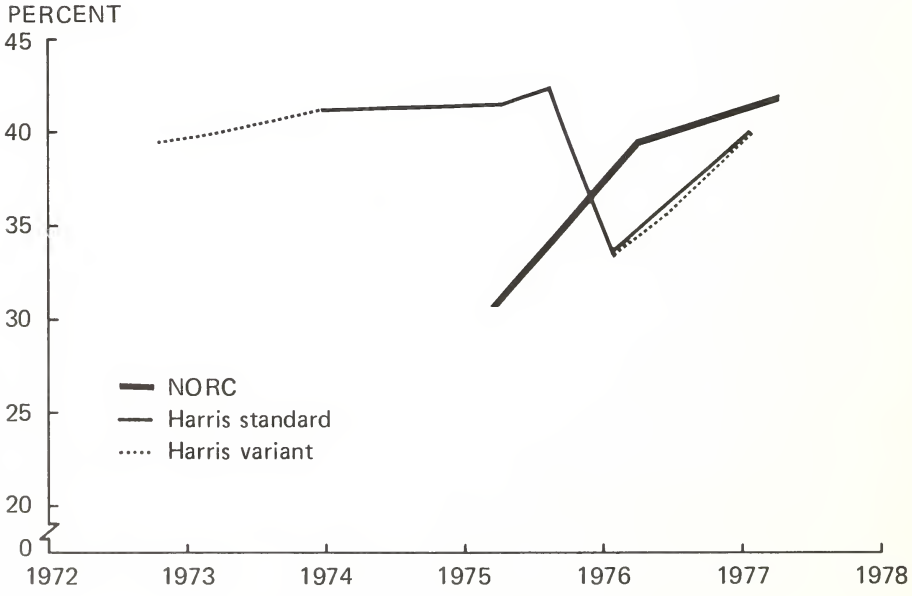


Figure 11. Great deal of confidence in banks and financial institutions.

The Harris trends on major companies (figure 10), are linear component declines, while the NORC trend is nonlinear. Both Harris series also report significantly less confidence than NORC (standard = −.042; variant = −.029).

On banks (figure 11), the Harris trends (1975-1977 only) are nonlinear, while NORC shows a strong linear increase. In their pooled proportions, NORC does not significantly vary from either Harris trend.

Summing up these comparisons, it appears that on the Congress, the executive branch, and labor, the NORC and Harris series show minimal divergence with similar trends and approximately the same level of confidence reported. On medicine, banks, and the press, there is some correspondence. Medicine has a similar direction to its trend but differences on the level of confidence, whereas banks and the press differ between houses on the trends but show similar levels of confidence. On science, the U.S. Supreme Court, organized religion, the military, and business, the houses show both different types of trends and moderate-to-large differences in their pooled proportions.

There is also evidence that there is some direction to the differences in confidence. On 8 of the 10 institutions on which the pooled proportions were compared, Harris registered lower mean confidence than NORC. Only on organized labor and financial institutions (for 1975-1977) did NORC register lower confidence than Harris. Across all 10 institutions, the average net difference was −.023 (Harris-NORC). Looking at this difference further, a similar comparison was made between NORC and Harris surveys done at approximately the same time (table 6). This comparison revealed the same eight-to-two split on institutions as the pooled averages and showed an average net difference of −.020. Much of the difference in direction disappeared, however, when Harris 2521 was excluded from the analysis. While 29 (.659) of the 44 comparisons at approximately the same time showed NORC getting more confident responses, only 19 (.559) of the 34 comparisons excluding Harris 2521 were more confident on NORC, and the net average fell to −.007. The exclusion of all the Harris 2521 comparisons overly compensates, of course, for its especially strong differences, but it does show that much of the directional difference originates from this source. In sum, there appears to be some tendency for Harris to find less confidence than NORC, but with the exception of Harris 2521, this difference is small.

Next, the rank-order association of the confidence items within and across houses was studied. Table 9 gives the ranking of institutions in question. The differing mixture of institutions on the various surveys hindered comparison, but two evaluations were made. First, on nine confidence items (major companies, organized labor, executive branch, Congress, the U.S. Supreme Court, organized religion, the press, the military, and medicine) common to all GSS and six Harris surveys, a comparison was made between the intrahouse rank-order correlations. On GSS, the Spearman's ρ between consecutive years were

Table 9. Rank Order of Institutions, by Survey

Rank	Harris 1574	Harris 1702	Harris 2131	Harris 2219	Harris 2236	Harris 2251	Harris 2319
1	MEDIC	MEDIC	MEDIC	FINAN	MEDIC	(NA)	MEDIC
2	FINAN	*EDUC*	EDUC	EDUC	FINAN	(NA)	FINAN
3	ARMY	*ARMY*	FINAN	FED	SCI	(NA)	BUS
4	EDUC	FINAN	SCI	BUS	ARMY	(NA)	CLERG
5	SCI	BUS	*ARMY*	TV	EDUC	(NA)	LABOR
6	BUS	SCI	*CLERG*	LABOR	CLERG	(NA)	(X)
7	JUDGE	LEGIS	BUS	(X)	JUDGE	(NA)	(X)
8	LEGIS	CLERG	*JUDGE*	(X)	FED	(NA)	(X)
9	*CLERG*	JUDGE	*FED*	(X)	BUS	(NA)	(X)
10	*FED*	FED	TV	(X)	LEGIS	(NA)	(X)
11	PRESS	PRESS	LEGIS	(X)	PRESS	(NA)	(X)
12	TV	TV	PRESS	(X)	TV	(NA)	(X)
13	LABOR	LABOR	LABOR	(X)	LABOR	(NA)	(X)

Rank	GSS 1973	Harris 2343	Harris 2354	NORC 4179	Harris 7482	GSS 1974	Harris 7487
1	MEDIC	MEDIC	MEDIC	JUDGE	MEDIC	MEDIC	MEDIC
2	EDUC	ARMY	SCI	CLERG	TV	EDUC	JUDGE
3	SCI	CLERG	FINAN	PRESS	BUS	SCI	ARMY
4	CLERG	JUDGE	CLERG	LEGIS	FED	CLERG	CLERG
5	ARMY	PRESS	PRESS	BUS	(X)	ARMY	FED
6	JUDGE	BUS	BUS	LABOR	(X)	JUDGE	PRESS
7	*BUS*	LEGIS	LEGIS	FED	(X)	BUS	BUS
8	*FED*	LABOR	LABOR	(X)	(X)	PRESS	LEGIS
9	LEGIS	FED	FED	(X)	(X)	TV	LABOR
10	PRESS	(X)	(X)	(X)	(X)	LABOR	(X)
11	TV	(X)	(X)	(X)	(X)	LEGIS	(X)
12	LABOR	(X)	(X)	(X)	(X)	FED	(X)
13	(X)	(X)	(X)	(X)	(X)	(X)	(X)

(See footnotes on p. 155.)

1973-1974 = .820
1974-1975 = .879
1975-1976 = .996
1976-1977 = .833
Average = .882

On the Harris surveys, the comparable figures were

2343-2434 = .783
2434-7581 = .993
7585-2521 = .967
2521-7690 = .900
Average = .880

These correlations suggest that there are no differences in the variability of institutional rankings between houses. Looking at the interhouse differences

Table 9. Rank Order of Institutions, by Survey—Continued

Rank	Harris 2430	Harris 2434	Harris 2515	GSS 1975	Harris 7581	Harris 7585	Harris 2521
1	MEDIC	MEDIC	EDUC	MEDIC	SCI	MEDIC	MEDIC
2	*JUDGE*	EDUC	TV	SCI	MEDIC	FINAN	FINAN
3	*TV*	JUDGE	ARMY	ARMY	FINAN	EDUC	CLERG
4	PRESS	CLERG	PRESS	FINAN	CLERG	TV	ARMY
5	FED	ARMY	BUS	EDUC	JUDGE	CLERG	JUDGE
6	LEGIS	TV	LABOR	JUDGE	PRESS	JUDGE	PRESS
7	BUS	PRESS	LEGIS	CLERG	ARMY	ARMY	BUS
8	(X)	LABOR	(X)	PRESS	BUS	PRESS	FED
9	(X)	FED	(X)	BUS	LEGIS	BUS	LABOR
10	(X)	LEGIS	(X)	TV	LABOR	LABOR	LEGIS
11	(X)	BUS	(X)	*LEGIS*	FED	FED	(X)
12	(X)	(X)	(X)	*FED*	(X)	LEGIS	(X)
13	(X)	(X)	(X)	LABOR	(X)	(X)	(X)

Rank	Harris 7681	GSS 1976	Harris 7684	Harris 2628	Harris 2630	Harris 7690	GSS 1977
1	JUDGE	MEDIC	JUDGE	TV	MEDIC	MEDIC	MEDIC
2	TV	SCI	FED	EDUC	SCI	FINAN	SCI
3	FED	FINAN	LEGIS	ARMY	FINAN	EDUC	FINAN
4	PRESS	ARMY	(X)	PRESS	TV	CLERG	CLERG
5	LEGIS	EDUC	(X)	BUS	PRESS	JUDGE	EDUC
6	BUS	JUDGE	(X)	LABOR	BUS	ARMY	ARMY
7	(X)	CLERG	(X)	LEGIS	FED	TV	JUDGE
8	(X)	PRESS	(X)	(X)	LEGIS	FED	FED
9	(X)	BUS	(X)	(X)	(X)	BUS	BUS
10	(X)	TV	(X)	(X)	(X)	PRESS	PRESS
11	(X)	LEGIS	(X)	(X)	(X)	LEGIS	LEGIS
12	(X)	FED	(X)	(X)	(X)	LABOR	TV
13	(X)	LABOR	(X)	(X)	(X)	(X)	LABOR

Note: Italics indicate ties in ranking.
NA Not available.
X Not applicable.

revealed the following correlations:

GSS 1973-Harris 2343 = .795
GSS 1974-Harris 2434 = .767
GSS 1975-Harris 7581 = .854
GSS 1976-Harris 2521 = .900
GSS 1977-Harris 7690 = .983
Average = .860

This is marginally smaller than the intrahouse surveys and since the time interval was shorter (an average of about 3 months between houses, 8 months for Harris and 12 months for GSS) it should have been higher. To look at the possibility that intersurvey differences in rankings were more variable than intrasurvey differences, a comparison was made between all surveys taken at

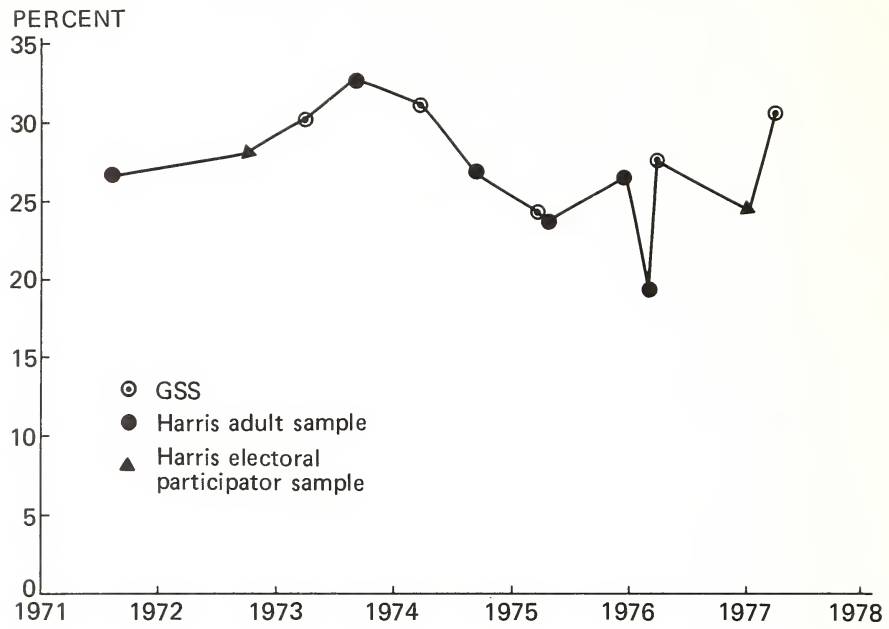


Figure 12. Nine-item confidence scale.

approximately the same time (table 6) and having at least six institutions in common. The four interhouse comparisons had an average ρ of .922 while the four intrahouse comparisons averaged only .862. Variability within houses over short time periods (contrary to the previous data) is as likely to be as great or greater than variability between houses.[3] In sum, the analysis shows a moderately high constancy in the rank ordering of institutions. This constancy is about the same for both houses and appears to be as strong between houses as it is within houses.

For a final comparison of the data, an analysis was made of a confidence scale. The nine-item scale was simply the average proportion responding a great deal of confidence on major companies, organized religion, the executive branch, organized labor, the press, medicine, the U.S. Supreme Court, Congress, and the military. Figure 12 shows the changes from 1971 to 1977 on the five GSS points and eight Harris points (six adult samples and two electoral participator samples). At several points, the series appear to agree quite closely. For example, GSS 1975 has a confidence score of .245 while Harris 7581 has .238. The biggest difference comes, as noted previously, between Harris 7585 (.267) and GSS 1976 (.276) on one hand and the intervening Harris 2521 (.196). With

[3] In the first case, the rankings were of the same list of institutions but over differing periods. In the second case, the time periods were much more similar but a differing number and mixture of institutions were ranked between the various surveys. For these and other reasons, neither case represents a perfect test of inter- versus intrasurvey variability in rankings.

the exception of this point, the combined series would seem to be largely in agreement, showing a rise in confidence from 1971 to 1973, a drop in confidence to early 1975, and a recovery generally prevailing until early 1977.

In conclusion, from the inspection of the differences in marginals and trends, both individual Harris and NORC surveys and the respective house series are often significantly variant. This variation appears across a number of intersurvey comparisons and trends but is distinctly highest for the 1976 divergence between Harris 2521 and GSS 1976. The rank-order comparisons indicate that differences in distributions and trends do not create larger shifts in rank between houses than within houses. Likewise, the analysis of the confidence scale, which could be expected to average over specific differences on particular institutions, shows a notable degree of compatibility between the Harris and NORC results. Therefore, there are sufficient particular differences to warrant some puzzlement and attempts at explanation, but the differences are limited in occurrence and magnitude.

ANALYSIS OF DIFFERENCES

From the preceding discussion it appears that there are some large differences between Harris and NORC marginals and trends on the confidence items (as well as some large intrahouse differences). Broadly speaking, there are four possible sources for these differences; (1) house effects, (2) survey effects, (3) item effects, or (4) true change. House effects are the result of differences in the standard organization and procedures of the different organizations. These effects include matters, which in general affect all surveys conducted by an organization, such as sample frame, survey method, and general interviewer training and instruction. Survey effects are the result of differences particular to the construction and operation of individual surveys. They include the population sampled, specific interviewer instructions, context and placement of the items, and the format and wording of the items. Item effects are caused by the nature and content of the items themselves, how they are understood and interpreted, and whether they are suitable and reliable measures. True changes are actual changes in evaluations of confidence after artifactual variations from house, survey, or item effects.

House Effects

In considering house effects, it was not possible to examine in detail every possible facet of the survey operations of Harris and NORC in order to document how they compared on each phase of operation and to assess the possible ramifications of differences. (For a detailed step-by-step comparison of survey procedures, see Bailar and Lanphier, 1977.) It was possible, however, to carry out some comparisons between Harris and NORC on the demographic profile of their samples, on the differences between block quota and full probability sampling, and on the handling of item nonresponse.

On Harris and NORC surveys, there were six demographics that were asked and coded in sufficiently similar fashions to permit interhouse comparisons: sex, age, education, family income, marital status, and religion. Comparison was made among GSS 1974, GSS 1975, and GSS 1976 and the five Harris surveys asked at approximately the same times, 7482, 2515, 7581, 2521, and 7681. On each of the demographics examined there appear to be consistant, small-to-moderate level differences in the sample populations. The GSS surveys averaged .454 male to Harris's .500 (d = .046); .393 over 50 years old to .358 (d = .035); .309 college educated to .379 (d = .070); .417 less than $10,000 family income to .440 (d = −.023); .185 widowed, separated, or divorced to .134 (d = −.051); and .644 Protestant to .610 (d = .034). It is not possible to determine whether these differences come from differences in the sample frame, the method of selecting respondents, nonresponse differentials, or other related factors.[4]

It is possible, however, to see what impact these differences might have on the reported confidence levels by standardizing the GSS surveys to match the Harris surveys. Since education both showed the largest disparity and was also related to more confidence items than the other demographics, the GSS surveys were weighted to match the education marginals on the Harris surveys. The impact of this standardization on the interhouse confidence differences was not great. On 15 confidence items appearing on the Harris and GSS surveys indicated above, there were no significant relationships between education and confidence, so standardization on education was unrelated to the interhouse differences. On the 18 items that showed a significant relationship between education and confidence, standardization increased the interhouse differences in 12 cases and decreased them in 6 instances. The net change over these 18 items was an increase in interhouse differences of .002 (or .001 over the 34 items). In brief, differences in the demographic profile of the samples have only negligible effects on confidence.

Next, a comparison of sample type was made. All NORC surveys used a multistage probability design to at least the block level. At that point, either a quota or full probability design was employed. In the quota approach, interviews must fill quotas for men under 35, men 35 and older, employed women, and unemployed women. These are filled by approaching households according to a fixed pattern and interviewing the first available people who fit the quota requirements. There is no enumeration of households or callbacks. In the full probability approach the eligible households have been prelisted, and selected households have been randomly chosen. These predesignated households are contacted and their members enumerated. A Kish table is then used to select the respondent. Repeated callbacks are made if needed to interview the designated respondent. No substitution of households or household members is

[4] These items are, of course, susceptible to differences due to other reasons beyond sample variation and house effects. For the sake of this comparison it is assumed that there are no significant response effects and so on.

allowed. The block quota design was used in GSS 1973, NORC 4179, and GSS 1974. GSS 1975 and GSS 1976 were experimental split ballots, half block quota and half full probability. GSS 1977 was a complete full probability survey.

The Harris surveys all use a multistage block quota design similar to the NORC block quotas except that the quota is only for sex rather than sex, age, and employment status.

By comparing the split halves on the 1975 and 1976 GSS, it was possible to determine whether sample type influenced confidence. Of the 26 comparisons (13 confidence items by 2 years), only one difference was statistically significant at the .05 level (and even this was not significant if corrected for clustering). Of course, this result only shows that the NORC full probability and block quota designs do not produce different results and does not directly indicate whether the Harris quota type might produce different results from either the NORC full probability or block quota sample approaches. It does, however, provide some basis for believing that sample type is not a likely source for large differences in attitude marginals. (For a more extensive discussion of the differences between full probability and block quota, see Stephenson, 1979.)

Last, the impact of differing interviewer training and instruction was partly assessed by looking at the handling of item nonresponse. An inspection of the Harris and NORC series revealed that Harris items had a consistantly higher level of no opinion responses than NORC did. Subtracting the average proportion replying with no opinion on the NORC surveys from those on the standard Harris series revealed the following surpluses on nonresponse:

CONBUS	= 0	CONJUDGE	= .014
CONCLERG	= .028	CONSCI	= .010
CONLABOR	= .016	CONLEGIS	= .009
CONPRESS	= .010	CONARMY	= .022
CONMEDIC	= .012	CONFINAN	= .014
CONFED	= .026	Average	= .015

This difference results from a house difference in interviewing instructions. NORC interviewers are instructed to probe for response while Harris interviewers apparently are not (see Smith, 1979; Schuman and Presser, 1978; Converse, 1976-1977; Louis Harris, n.d.). Since Harris items pick up more no opinion responses than NORC, they naturally pick up less responses in the three substantive evaluation categories. Eliminating no opinion responses from the analysis would therefore increase the proportion giving a great deal of confidence (and the other two substantive categories as well) more for Harris than for NORC. This would reduce in turn the difference between Harris and NORC whenever the NORC item had shown more confidence than Harris with the no opinion category included in the analysis. Since NORC did show more con-

fidence in 8 out of 11 instances this means that the exclusion of no opinion from analysis reduces the overall average difference between the houses or, to think of it another way, part of the differences between Harris and NORC are explained by house differences on no opinion. The reduction is of course not large. For example, on the executive branch, organized religion, and medicine, the pooled difference in the proportion with a great deal of confidence declines from .009 to .005, .029 to .020, and .043 to .037, respectively.

In sum, from the limited range of available information on house effects it was not possible to isolate any major source for interhouse differences. The largest source of differences appears to come from the handling of item nonresponse, and this apparently explains some of the differences in marginals.

Survey Effects

Next, considering survey effects, it was possible to examine differences due to (1) institutional descriptors, (2) external context (i.e., placement in questionnaire), and (3) internal context (ordering of institutions within the question). No evidence was available to study the impact of the several variations in question wording and format.

In July 1975, Proctor & Gamble conducted a random-digit-dialing telephone interview with a national adult sample of 364. This sample was split into three subsamples, and each was read a confidence question with a different set of institutional descriptors. The question asked:

> I'm going to name several institutions and groups in our country and for each of them I would like you to tell me whether you have a great deal of confidence, a moderate amount of confidence, or no confidence in it. For example, the first is _____. Would you say that you have a great deal of confidence, moderate amount of confidence, or no confidence in _____?

The different institutional descriptors used and the proportion responding "a great deal" are given in table 10.

On 8 of the 15 groups of institutional descriptors there was statistically significant variation in the proportion reporting a great deal of confidence. The term government enlists more support than either politicians or politics; U.S. President finds more confidence than Federal Government. Organized labor rates more confidence than big labor; the U.S. Supreme Court outranks judicial system or lawyers; established religion tops either organized religion or ministers and other religious leaders. The Army, Navy, and Air Force ranks first, the military second, and military leaders third; colleges bests educational system or professors; and automobile manufacturers outscores automobile dealers or automobile salesmen. In brief, by dressing up the different institutions in more or less flattering appellations (organized labor versus big labor), by focusing on

Table 10. Proportion With a Great Deal of Confidence, by Institutional Descriptor

Descriptor	Proportion	Descriptor	Proportion
Business leaders	.18	Doctors	.52
Business	.20	Hospitals	.49
Big business	.12	Medicine	.60
Politicians	.02	Ministers and other religious leaders	.35
Government	.20	Organized religion	.35
Politics	.04	Established religion	.50
U.S. Presidency	.30	The military	.48
Executive branch of Federal Government	.18	The Army, Navy, and Air Force	.63
Federal Government	.16	Military leaders	.21
Big government[a]	.06	Advertising	.16
Elected government officials[a]	.05	Advertising agencies	.10
Congress[a]	.07	Advertisers	.08
Organized labor	.21	Educational system	.32
Big labor	.07	Colleges	.46
Union leaders	.12	Professors	.29
U.S. Supreme Court	.35	Public opinion polls	.16
Judges[b]	.25	Election polls	.20
Judicial system	.15	Public opinion pollsters	.14
Lawyers[b]	.22	Automobile manufacturers	.14
Television news	.31	Automobile dealers	.04
Television news commentators	.23	Automobile salesmen	.05
Network television news	.25		
Newspapers	.19		
Newspaper publishers	.13		
The press	.13		

[a]Unlike most of the split items, these cover fairly different areas.
[b]Both judes and lawyers were asked on the same subsample.

different parts of an institution (U.S. Supreme Court versus judicial system), or in general by using an institutional rather than a generalized personal reference (colleges versus professors or the military versus military leaders), the confidence levels can be changed significantly.

External Context

In 1976, the General Social Survey and Harris 2521, fielded at about the same time, showed large differences in the amount of confidence Americans had in people running a number of national institutions. In general the Harris survey showed a considerably more negative appraisal of the institutional leadership than the GSS revealed. Upon examination of the questionnaires, it was found that the confidence question was the first item on the GSS and the confidence questions on the Harris survey followed shortly after a six-point alienation index. This index consists of four negatively phrased agree-disagree statements about various elite/leadership groups (e.g., the people running the country, the

rich, people with power, and the people in Washington) and two negatively phrased agree-disagree statements about efficacy and participation (see table 11 for wordings). It was hypothesized that these six negative statements (with four about elite/leaders) might have created a negative context and resulted in lower levels of confidence on the confidence questions.

In order to test that hypothesis, a split-ballot context experiment was set up on the 1978 GSS. A randomly preselected half of the sample was asked the alienation questions immediately before the confidence questions and the other half of the sample were asked the confidence questions before the alienation questions. As table 12 shows, the results were underwhelming. Of the 13 institutions involved, only one, confidence in major companies, showed a significant difference between the split ballots. Without the alienation questions preceding the confidence questions, 26.4 percent reported a great deal of confidence, but with the alienation questions first, only 19.0 percent had a great deal of confidence, a loss of 7.4 percent. In no other instance were any of the differences significant. There was, however, a small but general tendency for the confidence questions that followed the alienation questions to show less confidence. Of the 13 items, 9 had less confidence while 4 showed more confidence after the alienation questions. The sum difference over all 13 items was 17.7 percent, or an average drop of confidence of 1.4 percent per item. In sum, it appears that alienation did have an impact on confidence but much smaller than anticipated and centered on one item, major companies.

To examine why the alienation had little impact, the marginals of the alienation questions were checked. If people were giving positive (or disagree) responses to the negative alienation statements, it might be reasonable to posit that the negative connotation of these questions was overcome by the positive responses of the public. Table 13 shows that this was not the case. On five of the six questions (and all four of the elite/leadership questions), a clear majority agreed with the negative propositions of the questions. Thus, the negative marginals reenforce rather than weaken the argument that alienation should have a negative effect on confidence.

Next, it was decided to check if the alienation and confidence items were associated with each other. If being negative on alienation and lack of confidence were unrelated, it could be argued that the negative form and responses

Table 11. Harris Alienation Scale

Now I want you to read some things some people have told us they have felt from time to time. Do you tend to feel or not (*read list*)?

A. The people running the country don't really care what happens to you.
B. The rich get richer and the poor get poorer.
C. What you think doesn't count very much anymore.
D. You're left out of things going on around you.
E. Most people with power try to take advantage of people like yourself.
F. The people in Washington, D.C. are out of touch with the rest of the country.

Table 12. Marginals of Confidence, by Context

Order	Variable	p^a	Alienation later	Alienation first	Difference (later-first)
1	CONBUS	.002	.264	.190	.074
2	CONCLERG	.687	.329	.309	.020
3	CONEDUC	.872	.294	.284	.010
4	CONFED	.918	.126	.133	−.007
5	CONLABOR	.890	.114	.117	−.003
6	CONPRESS	.060	.180	.228	−.048
7	CONMEDIC	.320	.472	.456	.016
8	CONTV	.236	.141	.139	.002
9	CONJUDGE	.678	.303	.285	.018
10	CONSCI	.135	.421	.369	.052
11	CONLEGIS	.757	.130	.136	−.006
12	CONARMY	.329	.314	.299	.015
13	CONFINAN	.382	.351	.317	.034

Note: Marginal differences on the alienation questions by context were also inspected and no significant differences were found.

[a]Don't know answers excluded. All differences except major companies also not significant at .05 with don't know answers included.

on the alienation questions would not be transferred to the confidence question. In table 14, however, we see that the items are correlated, in the hypothesized direction (high alienation with low confidence). Of the 78 correlations, the sign is negative in 76 cases. While many of the associations (11) are insignificant, there are also a number (11) of moderately strong associations of .25 or over. These moderate associations are clustered among the three political institutions (executive branch of the Federal Government, U.S. Supreme Court, and Congress) and major companies. It appears that alienation and confidence are associated in the expected direction, which also seems to indicate that a context effect might occur.

In brief, the form of the alienation questions, their marginal distributions, and the association between confidence and alienation all indicate a potential context effect. Our examination of the split-ballot marginals showed little

Table 13. Alienation Marginals for Sample Preceding Confidence

Item	Feel	Not feel	*N*
A. The people running the country don't really care what happens to you.	.535	.465	729
B. The rich get richer and the poor get poorer.	.754	.246	732
C. What you think doesn't count very much anymore.	.576	.424	721
D. You're left out of things going on around you.	.284	.716	733
E. Most people with power try to take advantage of people like yourself.	.562	.438	731
F. The people in Washington, D.C. are out of touch with the rest of the country.	.597	.403	710

Table 14. Correlation Between Confidence and Alienation When Alienation Precedes Confidence

Institution	A	B	C	D	E	F
Banks and financial institutions	−.145	−.167	−.111	−.105	−.202	−.177
Major companies	−.176	−.314	−.179	−.186	−.262	−.208
Organized religion	−.110	(−.060)	−.079	(−.042)	−.095	−.168
Education	−.194	−.147	−.173	−.122	−.166	−.242
Executive branch of the Federal Government	−.308	−.181	−.264	−.165	−.273	−.366
Organized labor	−.107	(−.029)	−.079	(.009)	(−.049)	−.198
Press	−.082	−.061	−.096	(−.039)	−.106	−.086
Medicine	−.133	−.081	−.071	(−.047)	−.102	−.149
Television	(−.056)	(−.023)	−.063	(.028)	(−.052)	−.117
U.S. Supreme Court	−.241	−.192	−.260	−.131	−.225	−.298
Scientific community	−.165	−.100	−.191	−.128	−.166	−.145
Congress	−.286	−.166	−.238	−.157	−.257	−.346
Military	−.127	−.100	−.073	−.073	−.073	−.073

Note: Negative signs indicate high alienation associated with low confidence. Correlations in parentheses not significant at .05 level.

difference (although minimally in the hypothesized direction) except for confidence in major companies. The question becomes then, Why business and not the others? Looking again at the correlation matrix in table 14, we see that major companies are one of the institutions most strongly associated with the alienation questions. This explanation could be useful except for the fact that the other institutions with moderately strong associations (the three political institutions) show virtually no marginal differences because of context. It therefore appears that another explanation must be sought. A plausible alternative is that major companies showed the context effect, while the others did not, because it was the first institution in the confidence question. As the confidence item nearest to the alienation questions, it may have been more influenced than the other items. It would be desirable if these interpretation could be buttressed by an association between the item order of the other institutions and their context shifts, but no apparent pattern emerges.

Finally, we extended the search for context effects by examining the correlation matrices between alienation and confidence for both question orders. In table 15, the associations between confidence and alienation are given for the split-ballot half on which alienation followed confidence. When compared to table 14, we see that the associations were uniformly higher when alienation came first than when it followed the confidence item. In 65 instances, the association became more negative (including the 2 cases in which positive associations decreased), and in the remaining 12 cases the associations remained the same or increased. Looking at just the four elite/leadership alienation items, the effect is even stronger, with 47 associations increasing and 5 staying stable or decreasing. On the average, each of the 78 associations increased by .04, and each of the elite/leadership associations with confidence rose by .05. Making a

simple additive scale of alienation and confidence shows that the correlation between the scales is .289 when confidence comes first and .444 when alienation comes first. It is therefore apparent that the association between alienation and confidence is influenced by context. Without knowing what the true association would be with no context effect operating (e.g., if they were separated by a couple dozen questions on an interview), it is difficult to specify in what way context is working.

It is not known whether: (1) the appearance of the alienation questions first strengthens the relationship, or (2) the appearance of the confidence items first weakens the relationship, or (3) both. As a working hypothesis, however, the following scenario is proposed. The alienation questions help to provide a frame of reference by which the confidence of institutions in general and political institutions in particular are evaluated. Armed with this focused frame, people give responses to the confidence items in line with this reference. People variously move confidences up or down according to how this frame influences their perspective. The net result is that marginals are changed little (except on the major companies item), since people are moving confidence both up and down to bring it into line, but the associations between alienation and confidence are raised because of the constraint that alienation exercises.

In terms of marginal shifts, the alienation questions exercised minimal impact except on major companies, where the proximity or context effect was greatest (in Harris 2430, 2434, and 7487 and NORC 4179, where similar alienation scales, appeared there was little impact). As a result, the appearance of the alienation items prior to the Harris confidence questions in 1976 cannot explain the large and general differences between Harris and GSS on confidence at that time. Furthermore, since this was conceived as a strong test for an external

Table 15. Correlation Between Confidence and Alienation When Alienation Follows Confidence

Institution	A	B	C	D	E	F
Banks and financial institutions	−.084	−.098	−.132	−.102	−.156	−.161
Major companies	−.175	−.189	−.168	−.178	−.254	−.118
Organized religion	−.084	(−.060)	(−.054)	−.078	(−.005)	−.113
Education	−.141	−.076	−.117	(.049)	(−.025)	−.168
Executive branch of the Federal Government	−.270	−.082	−.269	−.076	−.146	−.275
Organized labor	(−.050)	(−.043)	(−.053)	(.038)	(−.007)	−.137
Press	−.090	(−.042)	(−.062)	−.074	(−.052)	−.073
Medicine	−.102	(−.038)	−.116	−.101	−.100	(−.046)
Television	(−.018)	(−.037)	(.001)	(.037)	(−.026)	(−.057)
U.S. Supreme Court	−.221	(−.058)	−.147	−.112	−.152	−.184
Scientific community	−.109	−.118	−.135	−.185	−.131	(−.049)
Congress	−.227	−.104	−.241	−.107	−.166	−.289
Military	−.071	(−.038)	(−.040)	(.018)	(−.049)	−.103

Note: Negative signs indicate high alienation associated with low confidence. Correlations in parentheses not significant at .05 level.

context effect, it is not likely that major ordering effects are influencing confidence marginals on the 1976 surveys. From the comparison of the correlation matrices, however, it is clear that context does exert a general impact—in this case on associations although not on marginals. It seems that the appearance of the alienation items first constrains the confidence rankings and strengthens the relationship between the scales.

An Aside on Harris 2521

If the alienation questions do not explain the large difference between Harris 2521 and GSS 1976, then what does? First, let's reiterate that Harris 2521 differs not only from GSS 1976 but also from Harris 7681 (conducted about a month after Harris 2521) and from the pooled average of Harris items from 1973 to 1977 (see table 8 and figure 12). Harris 2521 varies from Harris 7681 (mean difference, −.074) and from the pooled averages (−.061) by almost as much as it differs from GSS 1976 (−.079). Yet showing that Harris 2521 is an outlier does not explain why. An inspection of news events during the period of February-March 1976 does not reveal any apparent explanation for a sharp across the board rise and/or fall in confidence. Examination of the question wording, format, and order of institutions shows the usual amount of variation from GSS and other Harris versions. These differences undoubtedly account for some variation in responses but not for the large and unidirectional differences observed. Looking at the level of nonresponse (which would have lowered confidence if it had been extraordinarily high) revealed that on 8 of 10 items nonresponse was above the average for other Harris surveys from 1973 to 1977. The net average difference of .010 on Harris 2515 was, however, too small to account for much of the difference. Also, since we did not exactly replicate the context of GSS 1976 and Harris 2521 on the 1978 GSS experiment, it is possible that a context effect was operating but we misdiagnosed its source. Perhaps the placement of confidence first on the GSS questionnaire had an impact, or perhaps the general context and content on Harris 2521 had an influence. Yet these possibilities do not seem especially viable.

Finally, it was decided to see how responses to other questions on Harris 2521 compared to those on other Harris surveys. Data were available to compare the responses on the six alienation items to those on six other Harris surveys from 1974 to 1977 and on the Harris standard presidential job-rating question to ratings immediately before and after. On the six alienation questions, the Harris survey showed a deviation from the normal Harris level similar to that detected on confidence. The six items averaged .045 more than on the other surveys and the four political items (excluding rich/poor and being left out) averaged .057 more. This pattern did not show up on the presidential job-rating question, however. President Ford's rating fell between lower ratings in January and higher ratings in March. Thus Harris 2521 did not uniformly register low confidence, alienation, and disapproval. This of course opens up more questions than it answers. One might hypothesize that alienation had a major context

effect in 1976 because alienation was much stronger then. This is plausible but unprovable. Furthermore, there arises the question of why alienation was so high. The Harris alienation data show that alienation scores on Harris 2521 represent a peak, with levels dropping back after that point. It is impossible to tell whether this is a real crest of alienation or just curiously out of line as the confidence items appear to be. The crest interpretation is somewhat challenged by the typical, even improving level of presidential popularity, but these two measures are probably not highly correlated. It appears that the reason for low confidence on Harris 2521 remains a mystery. The high alienation levels are probably related to the low confidence scores, but the causal connection is uncertain (e.g., Did a real peak in alienation cause it to exert a context effect on confidence? Did the general content of the survey or some other context effect influence both alienation and confidence? Was there a real crisis of leadership that directly influenced both confidence and alienation but not Ford's job rating?). The bottom line is that the low confidence level on Harris 2521 is not readily explainable but does deviate from expected levels.

Internal Context

Two indications of possible ordering effects emerge from the data. The GSS has used only two orderings of institutions, one for the 1973, 1974, and 1977 surveys and another for 1975 and 1976. (The order also varies in that banking appears in the 1975-1977 surveys but not in 1973-1974. Since this item appears last on the list of institutions, it does not influence the other items.) This switch resulted in the order changes shown in table 16.

Marginal changes between 1973-1974 and 1975-1976, when there were no order changes, were compared to those in 1974-1975 and 1976-1977 when the

Table 16. Changes in Item Order: 1973-1977

Variable	1973-1974, 1977	1975-1976	Order change	
			1974-1975	1976-1977
CONBUS	1	5	+4	−4
CONCLERG	2	6	+4	−4
CONEDUC	3	7	+4	−4
CONFED	4	1	−3	+3
CONLABOR	5	2	−3	+3
CONPRESS	6	3	−3	+3
CONMEDIC	7	8	+1	−1
CONTV	8	9	+1	−1
CONJUDGE	9	10	+1	−1
CONSCI	10	11	+1	−1
CONLEGIS	11	12	+1	−1
CONARMY	12	4	−8	+8
CONFINAN	[a]13	13	(X)	(X)

X Not applicable.
[a]1977 only.

switches occurred. Items that moved up or down three or four places were compared to items that changed only a single position (CONARMY and CONFINAN were not considered). The mean absolute change in marginals between years for the items changing three or four positions were divided by the mean absolute change for items switching only a single position.

Ratio	1973-1974	1974-1975	1975-1976	1976-1977
±4/±1	1.41	2.88	2.03	2.57
±3/±1	1.27	0.60	0.72	3.04

This table shows that in years that a change in order occurred (1974-1975 and 1976-1977), the ratios varied more from unity than in the years that no ordering changes occurred. This suggests that part of the changes that were observed between 1974-1975 and 1976-1977 were due to the switches in ordering. It further suggests that ordering differences explain some of the difference in marginals between surveys.

Another indication of an ordering effect comes from work in progress by D. Garth Taylor. He has found that the confidence level of institutions is influenced by the confidence levels of the immediately preceding institution. If the first institution has a favorable ranking, it increases the confidence recorded on the following item, or if the first item has low confidence, it decreases the recorded confidence in the second item.

Among the several possible survey effects examined, it appears that institutional descriptors can sometimes influence the marginal evaluations (but since this factor was isolated in the preceding analysis of distributions and trends, it does not explain the differences that were still observed). External context was found in an experimental test to have a small marginal impact and a more noteworthy influence on correlations. Internal order showed indications of influencing marginals, but it was not possible to specify the precise manner or magnitude of its influence. Other unexamined variations in wording and format may also have contributed to differences in distributions and trends. Thus, while no single factor seems to be a major cause of differences, most appear to have some influence on confidence. Therefore, the multiple differences in the placement and construction of the confidence question probably added to the variation in responses between surveys to a notable but unspecified degree.

Item Effects

Several experiments designed to analyze the confidence question were conducted in the 1978 GSS. First, they examined how respondents interpreted and understood the confidence question. Respondents were asked how they defined confidence and what references they had in mind when they evaluated several specific institutions. Second, the association between differences in definitions

and references and confidence were examined. Third, through a postinterview reevaluation of responses to the confidence questions and a test/retest measure of attitude change, the reliability/stability of the confidence items was inspected.

What Does Confidence Mean?

On the 1978 General Social Survey, a random subsample participated in postinterview debriefing on the confidence questions. They were asked two questions about the meaning of the concept of confidence:

> When we ask about "confidence" in these questions, what does that word mean to you?
>
> Is there a word or phrase that would be more clear than "confidence" but would describe the same idea?

The object of these questions was to see if respondents understood the word confidence and to find out how they defined it. Table 17 groups their responses into 12 major categories. Approximately 95 percent of the respondents were able to give reasonable definitions of confidence. Only 2.2 percent declined to offer a definition—about the level giving a don't know reply to a typical attitude item, and another 3 percent gave a response that could not be considered a reasonable definition, most commonly consisting of attempts to define confidentiality. Of the 95 percent giving appropriate definitions, the overall favorite choice was that confidence in the people running institutions means trusting them. Almost 35 percent mentioned trust in their responses. In addition, the closely related terms having faith or believing in the leaders were selected by 10 percent and 12 percent, respectively. Also fairly closely related

Table 17. Respondents' Definitions of Confidence

Key word	Proportion
Trust	.345
Capability	.159
Believe in	.124
Faith	.100
Miscellaneous	.054
Honesty	.043
Common good	.037
Dependability	.034
Approval	.030
Incorrect response	.030
Sure	.022
Don't know, nothing	.022

Note: 830 responses, 738 cases.

to the idea of trust were 4 percent mentioning honesty, truthfulness, or some related term and the 2 percent replying that it meant you could be sure or certain of the leaders.

Another major emphasis in the definitions was capability. Almost 16 percent stated that having confidence in the people running institutions meant thinking that the leaders were competent and had the intellectual and practical abilities needed to carry out their duties.

Responses also related to this notion, as well as to the trust dimension, were those emphasizing dependability. This 3 percent tended to blend together the trust and capability dimensions and considered these two features to be part of dependability.

A third major distinction was made by the 3 percent that mentioned the common good. They stated that having confidence meant knowing the leaders were acting in the best interest of the country, that they were doing what the common welfare required rather than following either the wishes of special interests or their own personal inclinations. The final major distinction was in sharp contrast to the common good concept. This group (3 percent) stated that having confidence meant that the leadership was doing things that the respondent approved of, that they were carrying out policies that the respondent personally favored.

These different emphases were not mutually exclusive, however. Multiple responses were given by $12\frac{1}{2}$ percent of respondents. For all categories except miscellaneous and dependability, trust was the category most commonly accompanying other choices. For example, of the people mentioning capability, 30 percent also mentioned some other concept, with 10 percent of them also using the word trust. Similarly, of those choosing the common good, 42 percent also included another category, with trust leading again. Of the four major dimensions, only the common good and approval did not overlap at all.

When asked for a substitute term for confidence, the majority (58 percent) replied that there was no preferable word and that confidence was fully satisfactory. Those that offered alternatives gave the same list of terms they had mentioned previously, with 20 percent naming trust (or 48 percent of those mentioning an alternative); 4 percent, faith; $3\frac{1}{2}$ percent, believing in; 3 percent, dependability; 3 percent, honesty; 2 percent, capability; 1 percent, respect; and 1 percent, approval; and 5 percent miscellaneous and incorrect. Compared to the high level giving an acceptable definition to confidence (95 percent), the low level giving an alternative (42 percent) indicates that confidence is a meaningful and perhaps even preferred term for the evaluation of institutional leadership.

To the vast majority of people, confidence means trusting or having faith in the leadership, while a secondary group emphasizes competence and much smaller groups stress the concepts of service either to the common good or personal interests. In addition, a number of people gave definitions covering two or more

categories. It thus appears that confidence is a widely and correctly understood term, and while it has several different meanings associated with its use in the context of evaluating leaders, the concepts of trust and faith are central. These and the other meanings associated with confidence are close to the concepts included in the political trust/cynicism scale developed by the Center for Political Studies at the University of Michigan.

In addition, these differences in definition of confidence are not related to the level of confidence. Comparing the mean confidence score (from an additive scale on all 13 confidence items) to a series of dummy variables for each definition (trust, faith, believing in, honesty, certitude, dependability, capability, personal approval, common good, or incorrect definition) showed only one significant difference. Those defining confidence as personal approval were more confident than those not expressing this concept. In brief, while differences in definition exist, these differences are unrelated to the confidence level, and shifts in the definition of confidence (say, from trust to dependability) should have little impact on the confidence level.

Respondent References

As part of 1978 methodological experiments on confidence, a randomly selected half of the sample was asked who or what they were thinking about when particular institutions were mentioned. The questions covered the press, medicine, the scientific community, and the military and went as follows:

> Who do you think of when we ask you how much confidence you have in the people running _____? (Do you have any particular people or group in mind?)

Most people were able to come up with an organization, group of people, or individual, but a substantial minority could not offer a reference. On medicine, .897 gave a response but .012 gave responses that were irrelevant or misdirected. This left .885 with a relevant reference group. On the military, .863 gave responses but .011 had irrelevant responses, giving .852 with relevant responses. On the press, .817 gave answers but .023 had irrelevant answers, leaving .794 with relevant answers. On the scientific community, .657 gave a reference, but .052 gave irrelevant or wrong answers, including .023 who thought that scientific community meant their local community (place of residence), leaving .605 with relevant answers. While most people have some explicit reference in mind, a nontrivial minority of from 12 to 39 percent can offer no relevant reference point for their evaluation of confidence.

Next, we looked at what kind of references were given by those mentioning one. Answers were classified according to several different schemes. On all four institutions, answers were classified as personal or impersonal. Personal answers referred to people or groups of people. These were further broken down into those naming specific persons (e.g., Dr. Salk) and those naming groups of

people (e.g., doctors). Impersonal answers referred to organizations or topical subjects (e.g., hospitals or research). They were also subdivided into references to specific institutions (e.g., the Federal Drug Administration) or general groups and topics (e.g., medical schools or heart disease). All four items were also classified as referring to government or nongovernment bodies. Finally, each institution was subdivided into various categories relevant to the particular institution.

Table 18 compares the four institutions on the personal/impersonal and government/nongovernment variables. Personal references are highest for medicine (.688) followed by the press (.636), science (.459), and the military (.451). Few specific people are named in any of the areas, although including mentions of the President raises the military to .149. Specific references to groups and organizations are more common (except on the press) and on the military account for a plurality of references (.433). There are even larger differences among the institutions on references to the government. As would be expected, almost all military answers mentioned the government (.995). The government was also cited frequently for science (.307) and less frequently for medicine (.086) and the press (.020). From these comparisons, it is clear that people think of various institutions in different lights, emphasizing the impersonal and governmental in regard to the military, for example, and the personal and nongovernmental for medicine.

The institutions were also classified according to various schemes that disclosed some particular dimension within each institution. Table 19 shows that on science a near majority did not think of any substantive area. Space led, however, by a large margin over those areas that were mentioned, followed by medicine and more distantly by atomic energy and a wide scattering of other topics (electricity, chemistry, weather, etc.). On the press, three dichotomies were examined, media type (print versus electronic), geographic reference (local versus national or unspecified), and level of control (top management versus

Table 18. Personal/Impersonal and Government References, Selected Institutions

(Proportion)

Reference	Science (N = 506)	Press (N = 659)	Medicine (N = 756)	Military (N = 861)
Personal				
Specific	.026	.061	.017	[a].149
General	.433	.575	.671	.302
Impersonal				
Specific	.113	.053	.144	.433
General	.429	.311	.168	.116
Government	.307	.020	.086	.995

[a]Includes references to Carter, President Carter, or the President. If these were not counted as specific references to a person but as general references, the distribution would be .036, .415, .433, .116.

Table 19. Selected Institutional References

Reference	Marginal	N
Science		
Space	.252	
Medicine	.152	
Atomic	.028	508
Other	.106	
None	.463	
Press		
Electronic	.236	
Local	.051	590
Bosses (publishers, editors, etc.)	.176	
Medicine		
Doctors	.641	656
Research	.235	
Military		
Armed Forces	.679	
Civilian government	.308	
Private industry	.006	723
Other	.007	

others). Most people thought of the press in traditional terms as printed media, but almost one-fourth mentioned radio or television. National or unspecified citations also predominated. Selection of the top management was less common, only .176 of all choices and even among personal references only .277. More visible figures such as reporters and commentators were more commonly cited than their employers. On medicine, doctors were explicitly mentioned by .641, or about 93 percent of all personal references. Another common reference was to medical research, referred to by .235. On the military, responses were classified as mentioning the Armed Forces (.679), the civilian government (.308), private industry (.006), or other areas (.007). In brief, it appears that many different types of people, groups, and topics are thought of when people are asked to evaluate confidence in the major institutions.

The query then becomes whether these great differences in references lead to major differences in how much confidence people have in the various institutions. For example, do people who mention personal references have more (or less) confidence in the institution than people who make impersonal ones. Or do people who mention, say, the local press differ in their confidence rankings from those who do not. Table 20 shows the correlations between confidence in the four institutions and each of the reference categories cited above. Of the 37 relationships examined, there were significant differences in 10 instances. The basic pattern is that: (1) most differences were small, reference was not a major indicator of confidence level, and (2) specific or general personal references tended to associate with more confident ratings (differences in confidence ratings because of personal references are similar to differences from varying the institutional descriptors). Among the particular results, it was found that referring to space or no field was associated with confidence in science and that

Table 20. Correlation Between Confidence in Selected Institutions and Reference Categories

(Pearson's *r*)

Reference	Science	Press	Medicine	Military
Personal				
Specific	NS	.064	NS	.099
General	NS	NS	.122	.033
Impersonal				
Specific	NS	NS	−.070	.069
General	.108	NS	NS	NS
Government	NS	NS	NS	NS
Science:				
Space	.077	(X)	(X)	(X)
Medicine	NS	(X)	(X)	(X)
Atomic	NS	(X)	(X)	(X)
Other	NS	(X)	(X)	(X)
None	.074	(X)	(X)	(X)
Press:				
Electronic	(X)	NS	(X)	(X)
Local	(X)	NS	(X)	(X)
Bosses	(X)	NS	(X)	(X)
Medicine:				
Doctors	(X)	(X)	NS	(X)
Research	(X)	(X)	.180	(X)
Military:				
Armed Forces	(X)	(X)	(X)	NS
Civilian	(X)	(X)	(X)	NS
Private industry	(X)	(X)	(X)	NS
None, don't know	.143	NS	NS	NS

Note: All reference categories are coded as dichotomies. Positive value indicates that people mentioning the aspect were more confident.

NS Not significant at .05 level, not adjusted for multistage sampling.

X Not applicable.

referring to medical research was associated with confidence in medicine. The others showed no association. Overall it appears that one's frame of reference can influence one's confidence in an institution, but such an influence does not appear on many items and even when it does appear it is usually small.

Reliability/Stability

Next, an examination was made of the reliability/stability of the confidence items. As part of the postinterview evaluation of the confidence questions, respondents were handed back the questionnaire opened to the confidence question and asked to check over and change any answers they wanted to:

115. Now I am going to ask you a different kind of question. We would like you to help us understand more about the answers that people

give. We would also like you to help us understand what people really mean.

Turn to confidence item (page 23), fold back questionnaire and hand questionnaire and green pencil to R.

You might read over these questions to make sure that I marked the answers as you told me. Or maybe you would like to change your answers because you have had more time to think about questions.

Give respondent time to read questions and think about them. If made changes (Go to B). If made no changes (Ask A).

A. *If made no changes*: Are you sure that we have the answer you meant?

Yes (Go to next question)

No (Ask B)

B. *If made changes*: Did you have a second thought about the answers, didn't I get your answers right, or what?

.

.

.

122. Now that we talked a little more, would you like to change any of your answers to these questions or *add* anything to what I have already written down?

They were encouraged to reevaluate their responses, and at three points (115, 115A, and 122) they were asked if they wished to change their answers. Despite this encouragement, only from .011 to .022 of respondents changed their answers on any of the institutions (average = .016). Of all changes, .056 were from don't know to a substantive evaluation, .069 were from a substantive evaluation to don't know, .494 were in an upward direction, and .381 were in a downward direction. When asked why they had changed responses, the overwhelming majority (.92) said that they had changed their evaluations because they had had second thoughts or changed their minds; only .08 mentioned a misunderstanding, miscoding, or dissatisfaction with response categories. Thus, the vast majority of respondents gave a confidence rating that they were not willing to change even when encouraged to do so, and the changes that did occur represented mostly the vascillations of fence sitters rather than major problems with the measurement instrument.

In a further test of reliability/stability about 1 month after the initial interview, a one in five subsample was reinterviewed on the telephone and reasked several questions including the confidence items (for details, see Smith and Stephenson, 1979). An average of .633 of respondents gave the same substantive response both times. Dichotomizing responses into a great deal versus some and hardly any and a great deal and some versus hardly any and averaging the results over both collapsed and all 13 items gave an average agreement level of .805. This average was slightly lower than comparably dichotomized attitude items on

test/retest studies with the 1972, 1973, and 1974 GSS. These studies had average agreement levels of .846, .858, and .826. (Because of different intervals between test and retest, the 1972 rate would have been lower and the 1973 and 1974 rates higher had they had the same interval as in 1978.) It seems that the confidence items are subject to slightly more short term change than attitude items in general. Unfortunately these simple test/retest data do not permit distinction between changes caused by true alternations in attitudes (instability) and changes from inadequacies in the measurement instrument (unreliability). Some other evidence (the low proportions changing answers during postinterview debriefing, the indications that the questions were understood by respondents, and frequently large short term fluctuations in cross-sectional marginals) suggest that much of the change results from instability. Thus, the greater-than-average proportion changing responses on confidence compared to other attitude items may indicate that confidence is more subject to real short term fluctuations (instability) rather than to noise due to weaknesses in the measurement instrument.

Stability, Crystallization, and Conceptual Level

One reason for instability is that opinions are not crystallized, that is, that many people do not have firm opinions on the matter in question and their opinions represent only a leaning or nothing more than an almost random response to the question. Such uncrystallized opinions are susceptible to change either as the issue crystallizes and people assume a firmer position, which may or may not be the same as their uncrystallized responses, or as response effects or transitory real world stimuli influence the still uncrystallized opinions. Opinion can be uncrystallized for several reasons, such as lack of factual information, the newness of the issue, crosscutting pressures, low salience, or abstractness. The confidence questions do not appear to be especially troubled by the matters of information, newness, or crosscutting pressures. On saliency, the little information available suggests that the items are typical. Don't know answers are an indicator of uncrystallized opinion in general and low salience in particular (among other things, which naturally makes them a far from perfect measure of salience). The levels on the confidence questions on the GSS surveys range from .013 on medicine to .100 on the scientific community and average .036 for the 13 institutions. This average is typical for attitude items on the GSS, although the .100 responding don't know on the scientific community is distinctly higher than most attitude items and other confidence items. Also, on a measure of indirect salience on the 1978 GSS, two confidence items were included. The question asked:

> How often would you say that you and your friends think about the topics we've been discussing during the interview? Would you say that you and your friends think about (*Read each item A-E*) very often, sometimes, or almost never?

A. Women's rights
B. The people running organized labor
C. Satisfaction with their present financial situation
D. Laws about abortions
E. The scientific community

The most salient topic was personal finances (very often + sometimes = .798) followed by women's rights (.673), organized labor (.609), abortions (.522), and the scientific community (.465). Organized labor, a confidence item with probably middling salience, ranks about average, and the scientific community, probably the confidence item with the lowest salience (as the don't know answers also indicate) ranks fifth. In brief, the confidence items do not appear to suffer more from lack of salience than other typical attitude items do.

The confidence items do, however, probably have a higher degree of abstraction than many other attitude items. This can make it harder for items to become crystallized and as a result make changes in responses easier and more common. While people form conscious opinions to a certain extent on such matters as preferred presidential choice, position on capital punishment, or support for wage and price controls, they are probably less likely to have previously formulated positions on confidence in the Congress, the press, the scientific community, or other institutions.

Of course, to a greater or lesser extent people have some predispositions about different institutions (e.g., Congress is run by a bunch of crooks; doctors perform miracles; or big business and big labor don't care about the average citizen), but these do not represent a consciously preformulated opinion in the way that a candidate choice or position on a specific public issue does. On many opinion and political questions, respondents immediately identify the question as one they have thought about (e.g., If the presidential election were being held today, which candidate would you vote for—Humphrey, the Democrat; Nixon, the Republican; or Wallace, the candidate of the American Independent Party? or Do you favor or oppose the death penalty for persons convicted of murder?) and give responses that reflect their preconceived positions. On confidence, however, respondents have certain predispositions about institutions, but they do not have preexisting opinions that closely correspond to the query, Would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them? In other words, respondents cannot simply call up conscious, previously formulated responses and clearly (almost mechanically) use them to answer the question. Instead they must take a series of partly relevant predispositions and use these to respond to the structure of the confidence evaluation offered by the question.

The distinction between questions on which respondents have preexisting opinions that clearly supply answers to survey questions and those that do not is not absolute but a matter of degree. Still, it seems that the confidence questions lean toward the latter and are more likely to suffer from instability

due to summarizing and coding of the respondents' predispositions than in the former case.

Lack of crystallization does not, however, reflect on the technical adequacy of confidence as a measure instrument. It is not a function of the question being vague or ambiguous or having inappropriate response categories but a reflection of the abstract concept that one attempts to measure. Attitudes about confidence are not usually consciously preformulated in a summary and coherent fashion and cannot be simply or automatically plugged into any scale of responses. In essence, the nature of the topic of confidence in institutions probably helps to keep many attitudes uncrystallized and thus makes them more susceptible to changes.

In brief, it was found that there was variation in how confidence was defined and in the references cited, but these differences in focus had only weak and scattered influence on confidence ratings. The confidence items were found, however, to have a fairly high level of changeability across short time periods. Although definitive evidence is lacking, instability in attitudes rather than the unreliability of the item appears more likely to be the cause. This instability in turn seems in part due to the conceptual nature of the question and the problem of coding such attitudes.

True Change

No big single effect contributing to intersurvey differences seems to prevail, but several effects seem to contribute to the variation in distributions and trends. Another major cause for the observed differences is probably true change. The confidence items (or at least some of them) appear to be susceptible to episodic change. That is, episodes or events in the real world can cause rapid and sizable changes in the distribution of confidence (for other items of a similar episodic nature, see Smith, 1980; Mueller, 1973; Kernell, 1978; and Stimson, 1976). For example, on confidence in the executive branch of the Federal Government, Harris differed from NORC in part because it caught points before the Watergate disclosures and at Nixon's resignation that were missed by the NORC series. Also, the drop in confidence from .283 after Nixon's resignation but before his pardon by Ford to .200 and .177 after Ford's pardon of Nixon clearly shows the impact of an event (the pardon) on confidence (down 10 percent). The episodic nature of the confidence items probably results mainly from the fact that they evaluate the performance of particular leaders and groups (e.g., the President, the military, and organized labor), and this performance is subject to occasionally abrupt and well-publicized changes (e.g., the Nixon pardon or the Camp David accords). While many attitude items are not likely to be greatly influenced by a dramatic event, the confidence items are open to such episodic influence. This, is not, of course, an artifact but merely reflects a basic attribute of the measure. In addition, the uncrystallized state of many evaluations probably makes the item susceptible to even more changes

due to events than would otherwise be the case. In sum, while the confidence items may detect general consistencies and/or trends in the level of public confidence in particular institutions, they almost certainly also catch many short term episodic changes to events and conditions.

Based on the preceding analysis, it appears that no single prime cause can be found for the differences in distributions and trends between the Harris and NORC series. Many small effects do appear to be at work, however, including the handling of item nonresponse, institutional descriptors, and external and especially internal context and ordering. These and possibly other unexamined effects create or magnify differences between surveys. In addition, confidence is susceptible to sizeable short term shifts in marginals. The combination of the various response effects with the intrinsic instability of the measure makes a naturally bouncy item even more bouncy.

CONFIDENCE AS A SOCIAL INDICATOR

When one thinks of social indicators, one usually thinks of such demographics as the percent with a college education, per capita income, fertility rate, or per student educational expenditures. One then follows trends in these indicators to measure such specific changes as the educational upgrading of the labor force or more general changes such as the quality of life. More adventurously, one thinks of attitudinal social indicators such as the percent for capital punishment, the mean amount of anomia, level of political cynicism, or confidence in institutions. Again, as in the case of the demographics, the usual objective is to map the general trends in these measures or (as a typically less preferred alternative) to document stability.

This approach creates several problems. Many demographic indicators are subject to relatively little year-to-year fluctuation because the attributes they measure are not given to major variation over a wide range of normal conditions (e.g., barring some event like a world war). In addition, they are usually calculated with such care and from such large samples (e.g., the Current Population Survey) or from records of the total universe of events (vital statistics) that they are technically highly reliable and subject to a minimum of random variation. The attitudinal social indicators suffer by comparison in several ways. First, they typically have not been developed and tested as well as the methods used to measure the demographics, and they usually rely on sample bases that allow considerably more random variation than in the case of the demographic indicators. Second, less is known about how various response effects such as context or real world attributes (e.g., seasons) influence attitudes. Third, the attitudinal social indicators are subject to much greater real short term fluctuations than most demographics. Some of the fluctuation is due to the shift of uncrystallized attitudes and some to alternations of firmly held attitudes. (The distinction is one of degree but worth making because little and unimportant changes in the real world can shift uncrystallized attitudes but

firmly held and organized attitudes are moved only by larger and more notable events.)

Most of the time the greater sample variation, lesser development of the measurement instrument, possible response effects, and greater propensity for short term fluctuations does not create problems. And indeed substantively clear and technically adequate series exist that measure such matters as race relations, willingness to vote for a woman for President, political cynicism, and many other matters.

In other instances, such as confidence, good fortune runs out and problems in the form of significant differences in marginals and divergent trends between houses occur.[5] It has been shown that some of the variations are explained by a number of small but cumulatively important effects. Also, some are accounted for by especially large and not fully explainable differences between particular surveys. Additionally, much of the short term differences and perhaps most of the differences in trends are due to the real fluctuating nature of confidence. Thus, when we compared the Harris and NORC trends, we were not finding different trends so much as different points of bounce. Except for the extra variation created by the factors mentioned above (which of course complicates the interpretation of real trends and creates some major outliers), much of the inter- and intrasurvey changes in the trends are true fluctuations and not mostly artifactual aberrations.[6]

The problem is thus twofold. First, the many differences in survey procedures, format, wording, placement, and order add noise that hampers the accurate and consistent measurement of the real level of confidence and increases artificially the variation of responses. For results of greater reliability and precision, such differences have to be either eliminated from the confidence series, isolated, or adjusted for. The second problem is that confidence is not like a demographic social indicator nor can it be expected to act like one. The fundamental nature of the concept that is measured leads to uncrystallized and therefore unstable opinions, and its episodic nature further contributes to short term fluctuations. If one takes into account these two considerations (and their limitations), one can use the confidence items as measures of the fluctuating state of trust in major institutions.

---

[5] A general comparison of marginals from different houses revealed much smaller differences than were typical on confidence (Smith, 1979).

[6] Davis's analysis (1978) of trends in GSS items found that confidence had 3 of the 10 most variable items. Executive branch was the most variable of all items, organized religion was third, and education was fifth. Thus, confidence is highly variable even within the GSS series.

APPENDIX

Table A. Major Companies

Survey	Confidence (percent)					Respondents
	A great deal	Some	Hardly any	None	Don't know	
Harris 1702	46.6	31.7	4.4	1.6	15.8	1,026
Harris 2219	30.5	43.0	13.9	(X)	12.6	3,151
Harris 2236	26.8	44.3	16.2	4.3	8.5	1,596
Harris 2319	33.8	47.1	12.0	2.6	4.5	2,991
GSS 1973	29.3	53.3	10.8	(X)	6.7	1,500
Harris 2343	29.8	43.7	19.7	(X)	6.8	1,592
Harris 2354	27.6	52.0	16.1	(X)	4.3	1,482
NORC 4179	21.8	52.1	20.1	(X)	6.0	1,484
Harris 7482	24.1	51.4	19.6	(X)	4.8	1,476
GSS 1974	31.4	50.6	14.5	(X)	3.6	1,483
Harris 7487	21.7	52.2	20.8	(X)	5.4	1,518
Harris 2430	15.2	50.0	31.2	(X)	3.6	612
Harris 2434	15.9	48.4	32.5	(X)	3.2	1,522
GSS 1975	19.3	54.0	21.2	(X)	5.5	1,483
Harris 2515	18.1	48.0	28.0	(X)	5.9	1,836
Harris 7581	19.7	49.7	25.4	(X)	5.1	1,578
Harris 7585	19.7	48.2	25.2	(X)	7.0	1,491
Harris 2521	16.3	55.0	24.6	(X)	4.1	1,495
Harris 7681	21.5	52.4	21.9	(X)	4.2	1,519
GSS 1976	22.0	51.2	21.7	(X)	5.0	1,491
Harris 2628	20.5	50.1	23.2	(X)	6.2	1,801
Harris 2630	19.9	47.8	23.9	(X)	8.5	1,538
Harris 7690	20.4	51.2	23.0	(X)	5.4	1,519
GSS 1977	27.2	56.5	12.3	(X)	4.0	1,526
GSS 1978	21.6	57.9	16.0	(X)	4.4	1,529

X Not included in response categories.

Table B. Organized Religion

Survey	Confidence (percent)					Respondents
	A great deal	Some	Hardly any	None	Don't know	
Harris 1702	39.6	26.6	9.1	6.2	18.6	1,024
Harris 2236	29.4	39.5	14.9	6.8	9.4	1,596
Harris 2319	33.2	38.2	14.7	7.1	6.7	2,986
GSS 1973	34.8	45.8	15.9	(X)	3.5	1,495
Harris 2343	35.6	35.2	22.6	(X)	6.6	1,592
Harris 2354	28.9	41.8	22.1	(X)	7.2	1,481
NORC 4179	32.1	44.2	19.1	(X)	4.7	1,485
GSS 1974	44.3	42.8	10.8	(X)	2.1	1,481
Harris 7487	31.8	38.2	23.8	(X)	6.2	1,518
Harris 2434	32.0	40.0	22.2	(X)	5.9	1,521
GSS 1975	24.4	47.9	21.3	(X)	6.4	1,485
Harris 7581	32.2	39.0	20.2	(X)	8.5	1,576
Harris 7585	35.5	38.1	19.7	(X)	6.6	1,489
Harris 2521	23.7	42.4	24.4	(X)	9.4	1,494
GSS 1976	30.7	44.7	18.3	(X)	6.3	1,491
Harris 7690	29.3	40.9	22.3	(X)	7.5	1,519
GSS 1977	40.0	45.1	11.6	(X)	3.3	1,526
GSS 1978	30.7	47.3	18.2	(X)	3.8	1,526

X Not included in response categories.

Table C. Education

Survey	Confidence (percent)					Respondents
	A great deal	Some	Hardly any	None	Don't know	
Harris 1702	55.5	29.1	4.9	1.4	9.1	1,021
Harris 2219	31.0	38.0	14.2	(X)	16.8	3,146
Harris 2236	33.4	46.3	13.1	3.1	4.1	1,593
GSS 1973	37.0	53.4	8.2	(X)	1.4	1,495
Harris 2343	44.2	37.5	14.4	(X)	3.8	1,594
Harris 2354	45.5	40.5	10.4	(X)	3.5	1,480
GSS 1974	49.1	41.4	8.2	(X)	1.4	1,480
Harris 7481	39.1	44.8	12.4	(X)	3.7	1,515
Harris 2434	39.3	44.3	13.7	(X)	2.6	1,520
Harris 2515	36.2	43.6	15.1	(X)	5.1	1,833
GSS 1975	30.9	54.6	12.8	(X)	1.7	1,488
Harris 7581	36.1	42.6	17.3	(X)	3.9	1,574
Harris 7585	36.6	39.7	17.4	(X)	6.3	1,480
Harris 2521	27.9	51.8	17.1	(X)	3.3	1,493
GSS 1976	37.5	45.1	15.4	(X)	2.0	1,489
Harris 2628	31.7	46.1	17.0	(X)	5.2	1,797
Harris 7690	37.0	46.6	11.6	(X)	4.8	1,513
GSS 1977	40.6	49.6	8.8	(X)	0.9	1,526
GSS 1978	28.5	55.0	15.1	(X)	1.4	1,528

X Not included in response categories.

Table D. Executive Branch

Survey	Confidence (percent)					Respondents
	A great deal	Some	Hardly any	None	Don't know	
Harris 1702	37.2	38.7	9.0	3.1	12.0	1,025
Harris 2219	33.6	41.1	18.2	(X)	7.1	3,154
Harris 2236	27.2	47.0	14.9	3.2	7.7	1,589
GSS 1973	29.3	50.4	18.4	(X)	1.9	1,498
Harris 2343	19.4	39.7	34.3	(X)	6.7	1,590
Harris 2354	13.4	42.2	41.3	(X)	3.1	1,478
NORC 4179	14.2	48.9	33.0	(X)	3.8	1,483
Harris 7482	11.7	40.9	44.4	(X)	3.0	1,471
GSS 1974	13.6	42.5	41.7	(X)	2.2	1,482
Harris 7487	28.3	51.9	13.7	(X)	6.1	1,517
Harris 2430	20.0	55.2	20.6	(X)	4.3	611
Harris 2434	17.7	55.0	22.9	(X)	4.4	1,520
GSS 1975	13.3	54.6	29.5	(X)	2.6	1,488
Harris 7581	13.1	49.2	32.9	(X)	4.8	1,572
Harris 7585	16.0	48.2	28.7	(X)	7.0	1,478
Harris 2521	10.8	55.3	26.6	(X)	7.3	1,488
Harris 7681	16.5	54.8	23.7	(X)	5.1	1,517
GSS 1976	13.5	58.5	25.0	(X)	3.0	1,494
Harris 7684	22.3	51.0	23.4	(X)	3.2	1,438
Harris 2630	14.5	50.5	24.9	(X)	10.1	1,540
Harris 7690	23.3	55.6	14.2	(X)	6.9	1,515
GSS 1977	27.9	54.4	14.5	(X)	3.1	1,525
GSS 1978	12.5	59.4	24.9	(X)	3.2	1,528

X Not included in response categories.

Table E. Organized Labor

Survey	Confidence (percent)					Respondents
	A great deal	Some	Hardly any	None	Don't know	
Harris 1702	19.6	38.2	18.4	10.3	13.5	1,021
Harris 2219	10.3	34.3	38.9	(X)	16.6	3,151
Harris 2236	15.3	43.7	24.1	8.7	8.1	1,587
Harris 2319	22.9	44.2	19.0	8.3	5.6	2,993
GSS 1973	15.5	54.6	25.7	(X)	4.1	1,495
Harris 2343	19.8	40.2	32.7	(X)	7.3	1,591
Harris 2354	16.2	49.8	28.2	(X)	5.7	1,480
NORC 4179	18.7	49.7	27.4	(X)	4.2	1,484
GSS 1974	18.2	53.5	25.5	(X)	2.8	1,481
Harris 7487	17.4	52.3	24.1	(X)	6.2	1,508
Harris 2434	18.5	46.4	30.9	(X)	4.3	1,520
Harris 2515	16.3	40.8	33.6	(X)	9.3	1,826
GSS 1975	10.1	54.2	29.3	(X)	6.4	1,488
Harris 7581	13.5	40.7	37.2	(X)	8.6	1,571
Harris 7585	18.0	39.5	34.5	(X)	7.9	1,485
Harris 2521	9.9	46.5	35.7	(X)	7.8	1,489
GSS 1976	11.6	47.5	33.0	(X)	7.9	1,494
Harris 2628	10.6	38.1	42.5	(X)	8.8	1,797
Harris 7690	14.5	43.1	35.9	(X)	6.5	1,519
GSS 1977	14.8	49.7	31.7	(X)	3.9	1,524
GSS 1978	11.0	46.3	37.6	(X)	5.1	1,528

X Not included in response categories.

Table F. Press

Survey	Confidence (percent) A great deal	Some	Hardly any	None	Don't know	Respondents
Harris 1702	26.5	45.8	12.8	5.6	9.4	1,027
Harris 2219	16.5	49.2	22.0	(X)	12.3	3,147
Harris 2236	18.4	50.4	21.0	5.5	4.8	1,594
GSS 1973	23.1	60.7	14.7	(X)	1.5	1,500
Harris 2343	30.3	45.0	21.5	(X)	3.2	1,592
Harris 2354	27.8	53.1	16.7	(X)	2.4	1,481
NORC 4179	25.1	51.1	21.3	(X)	2.6	1,481
GSS 1974	25.9	55.4	17.5	(X)	1.2	1,481
Harris 7487	24.8	48.2	23.4	(X)	3.7	1,514
Harris 2430	30.9	45.8	21.6	(X)	1.6	611
Harris 2434	25.6	48.2	24.7	(X)	1.6	1,521
GSS 1975	23.9	55.5	17.9	(X)	2.8	1,484
Harris 7581	25.9	52.5	19.3	(X)	2.2	1,577
Harris 7585	27.5	47.0	20.8	(X)	4.7	1,482
Harris 2521	20.1	50.3	25.4	(X)	4.2	1,490
Harris 7681	21.3	52.0	24.6	(X)	2.0	1,518
GSS 1976	28.5	52.1	17.7	(X)	1.8	1,490
Harris 2628	25.0	51.9	19.6	(X)	3.4	1,798
Harris 2630	24.7	51.6	18.8	(X)	5.0	1,541
Harris 7690	17.8	55.1	23.3	(X)	3.8	1,515
GSS 1977	25.1	57.3	15.5	(X)	2.2	1,526
GSS 1978	20.1	58.4	19.7	(X)	1.8	1,528

X Not included in response categories.

Table G. Medicine

Survey	Confidence (percent) A great deal	Some	Hardly any	None	Don't know	Respondents
Harris 1702	60.5	27.3	4.1	1.1	7.0	1,023
Harris 2236	48.2	36.1	8.8	2.8	4.1	1,591
Harris 2319	62.9	26.8	6.4	2.0	1.9	1,991
GSS 1973	54.1	39.2	5.7	(X)	0.9	2,496
Harris 2343	57.6	30.8	9.6	(X)	2.0	1,591
Harris 2354	59.9	33.3	5.5	(X)	1.3	1,482
Harris 7482	52.6	35.9	9.8	(X)	1.6	1,472
GSS 1974	60.4	33.7	4.5	(X)	1.5	1,482
Harris 7487	49.3	40.1	7.6	(X)	3.0	1,512
Harris 2430	49.7	35.1	12.9	(X)	2.3	612
Harris 2434	48.5	38.1	11.7	(X)	1.7	1,518
GSS 1975	50.5	40.1	7.9	(X)	1.5	1,487
Harris 7581	42.8	41.7	11.5	(X)	4.1	1,576
Harris 7585	53.7	32.4	10.3	(X)	3.6	1,480
Harris 2521	42.0	43.0	11.7	(X)	3.4	1,492
GSS 1976	54.1	35.3	9.2	(X)	1.3	1,492
Harris 2630	50.1	34.3	9.7	(X)	5.9	1,543
Harris 7690	42.5	44.3	11.0	(X)	2.2	1,516
GSS 1977	51.5	41.2	6.2	(X)	1.1	1,526
GSS 1978	46.0	44.0	9.2	(X)	0.8	1,527

X Not included in response categories.

Table H. Television

Survey	Confidence (percent) A great deal	Some	Hardly any	None	Don't know	Respondents
Harris 1702	20.3	43.8	18.8	7.4	9.7	1,019
Harris 2219	15.7	43.9	21.3	(X)	19.1	3,143
Harris 2236	17.9	51.4	21.9	4.5	4.3	1,594
GSS 1973	18.6	58.5	21.8	(X)	1.1	1,497
Harris 2343	40.3	43.5	14.1	(X)	2.1	1,593
Harris 2354	36.6	50.2	11.7	(X)	1.6	1,481
Harris 7482	34.2	46.8	17.3	(X)	1.7	1,474
GSS 1974	23.4	58.1	17.3	(X)	1.1	1,481
Harris 2430	36.2	47.9	15.0	(X)	1.0	608
Harris 2434	32.3	50.0	16.5	(X)	1.2	1,519
Harris 2515	33.6	49.2	14.8	(X)	2.3	1,835
GSS 1975	17.8	57.4	22.4	(X)	2.4	1,486
Harris 7585	36.6	45.2	14.1	(X)	4.1	1,478
Harris 2521	27.9	51.8	17.1	(X)	3.3	1,493
Harris 7681	28.3	53.7	16.0	(X)	2.0	1,515
GSS 1976	18.7	52.3	27.2	(X)	1.7	1,490
Harris 2628	32.6	51.4	13.3	(X)	2.7	1,801
Harris 2630	34.5	45.6	15.4	(X)	4.5	1,541
Harris 7690	27.6	54.2	15.9	(X)	2.4	1,517
GSS 1977	17.4	55.9	25.1	(X)	1.5	1,525
GSS 1978	13.8	53.4	31.0	(X)	1.8	1,526

X Not included in response categories.

Table I. U.S. Supreme Court

Survey	Confidence (percent) A great deal	Some	Hardly any	None	Don't know	Respondents
Harris 1702	39.5	29.0	12.9	7.8	10.7	961
Harris 2236	28.5	42.3	15.6	5.8	7.7	1,594
GSS 1973	31.5	49.8	15.4	(X)	3.3	1,497
Harris 2343	33.3	40.0	20.2	(X)	6.5	1,591
NORC 4179	34.1	44.0	16.0	(X)	5.8	1,485
GSS 1974	33.2	47.9	14.4	(X)	4.5	1,482
Harris 7487	40.1	41.3	13.5	(X)	5.1	1,515
Harris 2430	34.8	44.9	15.6	(X)	4.8	610
Harris 2434	35.0	44.0	16.7	(X)	4.3	1,521
GSS 1975	30.8	46.3	18.6	(X)	4.3	1,485
Harris 7581	28.7	43.8	21.5	(X)	6.0	1,575
Harris 7585	27.5	41.8	21.2	(X)	9.4	1,482
Harris 2521	21.9	47.5	22.4	(X)	8.1	1,489
Harris 7681	31.6	43.3	20.9	(X)	4.3	1,519
GSS 1976	35.4	43.6	5.4	(X)	5.6	1,491
Harris 7684	37.9	39.8	18.5	(X)	3.8	1,435
Harris 7690	28.6	47.7	17.9	(X)	5.8	1,516
GSS 1977	35.7	49.4	10.8	(X)	4.1	1,522
GSS 1978	28.1	52.8	14.6	(X)	4.5	1,527

X Not included in response categories.

Table J. Scientific Community

Survey	Confidence (percent)					Respondents
	A great deal	Some	Hardly any	None	Don't know	
Harris 1702	45.1	27.9	5.3	1.5	20.3	951
Harris 2236	36.8	38.8	5.8	1.4	17.2	1,589
GSS 1973	36.9	47.1	6.5	(X)	9.5	1,495
Harris 2354	45.5	36.6	5.7	(X)	12.3	1,480
GSS 1974	45.0	37.7	6.7	(X)	10.6	1,481
GSS 1975	37.7	45.2	6.5	(X)	10.7	1,487
Harris 7581	47.9	34.6	7.8	(X)	9.7	1,572
GSS 1976	42.9	38.0	7.5	(X)	11.6	1,486
Harris 2630	44.4	37.9	7.9	(X)	9.8	1,538
GSS 1977	41.0	45.7	5.5	(X)	7.8	1,522
GSS 1978	36.2	48.3	7.3	(X)	8.3	1,527

X Not included in response categories.

Table K. Congress

Survey	Confidence (percent)					Respondents
	A great deal	Some	Hardly any	None	Don't know	
Harris 1702	40.9	41.5	6.9	2.1	8.6	1,021
Harris 2236	21.0	56.8	14.0	2.3	6.0	1,591
GSS 1973	23.5	59.0	14.9	(X)	2.6	1,497
Harris 2343	29.7	48.6	16.2	(X)	5.5	1,590
Harris 2354	17.1	55.6	24.0	(X)	3.3	1,479
NORC 4179	22.7	57.5	15.6	(X)	4.2	1,485
GSS 1974	17.1	59.0	20.9	(X)	3.0	1,481
Harris 7487	17.8	58.0	20.6	(X)	3.6	1,515
Harris 2430	16.2	63.2	19.0	(X)	1.6	611
Harris 2434	16.4	60.2	21.1	(X)	2.4	1,519
Harris 2515	12.4	49.2	34.5	(X)	3.9	1,837
GSS 1975	13.3	58.6	25.2	(X)	2.9	1,487
Harris 7581	13.6	51.7	30.4	(X)	4.3	1,576
Harris 7585	12.1	49.0	32.5	(X)	6.4	1,488
Harris 2521	8.8	52.2	33.3	(X)	5.6	1,491
Harris 7681	17.9	55.1	23.9	(X)	3.1	1,516
GSS 1976	13.7	58.2	25.5	(X)	2.6	1,494
Harris 7684	16.7	54.7	25.9	(X)	2.7	1,434
Harris 2628	9.5	48.8	36.9	(X)	4.8	1,801
Harris 2630	12.7	53.4	27.9	(X)	6.0	1,539
Harris 7690	16.5	54.4	25.0	(X)	4.0	1,518
GSS 1977	19.1	60.9	17.1	(X)	2.9	1,523
GSS 1978	12.9	63.1	20.9	(X)	3.1	1,527

X Not included in response categories.

Table L. Military

Survey	Confidence (percent) A great deal	Some	Hardly any	None	Don't know	Respondents
Harris 1702	55.5	29.2	3.4	1.7	10.1	1,016
Harris 2236	36.1	41.2	11.9	5.1	5.8	1,594
GSS 1973	31.7	49.5	16.1	(X)	2.7	1,498
Harris 2343	40.5	35.2	18.4	(X)	5.9	1,592
GSS 1974	39.6	44.4	13.4	(X)	2.6	1,483
Harris 7487	33.9	44.0	16.8	(X)	5.3	1,517
Harris 2434	30.7	43.6	21.4	(X)	4.3	1,522
Harris 2515	26.7	41.0	25.9	(X)	6.4	1,836
GSS 1975	35.2	45.8	14.3	(X)	4.6	1,487
Harris 7581	24.5	44.3	24.5	(X)	6.7	1,575
Harris 7585	30.3	41.9	20.9	(X)	6.9	1,480
Harris 2521	22.5	49.7	21.2	(X)	6.8	1,491
Harris 7681	36.2	44.1	15.3	(X)	4.3	1,520
GSS 1976	39.2	41.3	13.3	(X)	6.2	1,491
Harris 2628	30.4	40.0	22.5	(X)	7.1	1,800
Harris 7690	27.6	49.0	16.9	(X)	6.5	1,517
GSS 1977	36.3	50.3	10.3	(X)	3.1	1,526
GSS 1978	29.5	54.0	12.8	(X)	3.7	1,528

X Not included in response categories.

Table M. Banks and Financial Institutions

Survey	Confidence (percent) A great deal	Some	Hardly any	None	Don't know	Respondents
Harris 1702	54.3	31.5	3.8	1.4	9.1	1,023
Harris 2219	59.1	26.2	3.0	(X)	11.7	3,147
Harris 2236	39.1	44.3	8.0	2.8	5.8	1,590
Harris 2354	41.2	44.9	9.6	(X)	4.3	1,476
GSS 1975	31.9	54.0	11.1	(X)	3.0	1,488
Harris 7581	41.5	44.0	11.3	(X)	3.2	1,574
Harris 7585	42.3	42.3	10.4	(X)	4.9	1,481
Harris 2521	33.5	52.5	10.6	(X)	3.4	1,491
GSS 1976	39.5	48.1	10.0	(X)	2.4	1,492
Harris 2630	36.0	44.7	12.7	(X)	6.5	1,538
Harris 7690	40.0	46.8	10.0	(X)	6.5	1,513
GSS 1977	41.9	47.4	8.8	(X)	1.8	1,526
GSS 1978	32.9	54.0	11.7	(X)	1.4	1,528

X Not included in response categories.

REFERENCES

Bailar, Barbara, and C. Michael Lanphier. 1977. "Development of Survey Methods to Assess Survey Practices: A Report of the American Statistical Association Pilot Project on the Assessment of Survey Practices and Data Quality in Surveys of Human Population." Washington, D.C.: American Statistical Assocation.

Converse, Jean M. 1976-1977. "Predicting No Opinion in the Polls." *Public Opinion Quarterly* 40: 515-530.

Davis, James A. 1978. *Trends in NORC General Social Survey Items, 1972-1977.* GSS Technical Report No. 9. Chicago: National Opinion Research Center.

Davis, James A., Tom W. Smith, and C. Bruce Stephenson. 1978. *General Social Surveys, 1972-1978: Cumulative Codebook.* Chicago: National Opinion Research Center.

Kernell, Samuel. 1978. "Explaining Presidential Popularity . . ." *American Political Science Review* 76.

Ladd, Everett Carll, Jr. 1976-1977. "The Polls: The Question of Confidence." *Public Opinion Quarterly* 40: 544-552.

Louis Harris and Associates. No date. "About Interviewing." Instruction memorandum. Louis Harris Data Center, University of North Carolina.

Mueller, John E. 1973. *War, Presidents, and Public Opinion.* New York: John Wiley and Sons.

Procter & Gamble. 1975. "Highlights: Research Pilot Study—Public Opinion Polls." Unpublished paper.

____. 1975. "Research Exploration—Public Opinion Polls." Unpublished paper.

Santi, Lawrence. 1978. "Confidence in Selected Institutions in 1975: An Attempt at Replication Across Two National Surveys." Paper presented at the Annual Meeting of the Pacific Sociological Society, Spokane, Washington.

Schuman, Howard, and Stanley Presser. 1978. "The Assignment of 'No Opinion' in Attitude Surveys." In Karl F. Schuessler, ed., *Sociological Methodology, 1979.* San Francisco: Jossey-Bass.

Smith, Tom W. 1979. "In Search of House Effects: A Comparison of Responses to Various Questions by Different Survey Organizations." *Public Opinion Quarterly* 42: 443-463.

____. 1980. "America's Most Important Problem: A Trend Analysis, 1946-1976." *Public Opinion Quarterly* 44: 164-180.

Smith, Tom W., and D. Garth Taylor. 1980. "Public Opinion and Public Regard for the Federal Government." In Carol Weiss and Allen Barton, eds., *Making Bureaucracies Work.* Beverley Hills: Sage.

Stimson, James A. 1976. "Public Support for American Presidents: A Cynical Model." *Public Opinion Quarterly* 40.

Taylor, D. Garth. 1976. "Procedures for Evaluating Trends in Qualitative Indicators." In James A. Davis, ed., *Studies in Social Change Since 1948.* NORC Report No. 127a. Chicago: National Opinion Research Center.

Turner, Charles F., and Elissa Krauss. 1978. "Fallible Indicators of the Subject State of the Nation." *American Psychologist* 33: 456-470.

Afterword

Edwin D. Goldfield
National Research Council
National Academy of Sciences

The tools and trappings of survey research have increased and improved greatly in the last several decades. We have new collection procedures and devices such as randomized response, random-digit dialing, computer-controlled calling, and machine-readable forms; we have more sophisticated hardware and software for data processing; and we have more elaborate analytical and display techniques. As a respondent, I am a little dubious about some of the mechanical advances. Recently I was called on my home telephone by a recording that proceeded to interview me. More fundamental is the improvement in our knowledge of the mind-to-mind interaction in asking questions and answering them. Here progress is harder and slower, especially in the domain of subjective phenomena. But more attention is being paid and progress is being made. This monograph is an example.

The measurement of subjective phenomena is not standardized and fully definitive and may never be so, but it is established as an important arm of social research. An indication of rapid change in availability, acceptance, and use is found in the introductory texts of the Federal Government publications, *Social Indicators 1973* and *Social Indicators 1976*. The first said, "*Social Indicators 1973* is restricted almost entirely to data about objective conditions." The second said, "Three broad types of indicators may be distinguished in this report: indicators of system performance, indicators of well-being, and indicators of public perceptions. . . . Although all three types of indicators are important, this report focuses primarily on indicators of well-being and public perceptions."

With some lag relative to nongovernmental survey research, Government agencies are becoming collectors and users of subjective measures. The majority of the subjective indicators reported in *Social Indicators 1976* came from nonfederal sources, but the Federal role as collector or sponsor is increasing. As a former official of the U.S. Bureau of the Census, I can remember when that agency, with myself as one of the spokespersons, self-righteously proclaimed that its inquiries were limited to supposedly objective questions. The chapter by Donald Dahmann in this monograph indicates that such a constraint is no longer deemed necessary or desirable.

The distinction between objective and subjective measures was recognized early in the history of subjective indicators. But that distinction is becoming less marked. It is no longer common practice to label proposed questions as subjective as an argument for proscribing them, and we recognize more and more that so-called objective questions have subjective overtones. Age is an oft-cited example of a truly objective measure, but its definition and perception vary among different cultures and it is not always objectively reported in this country. The census questions on race and ethnicity are admittedly subjective for many respondents. Despite 40 years of effort, it has not been possible to eliminate subjective elements completely from the reporting of labor force status. Household relationship is another subjectively tinged example.

Without going so far as to claim all measures as subjective, we can nevertheless agree that a large portion of the spectrum of inquiries is primarily or entirely subjective. The study of the measurement of subjective phenomena is important because it covers so much and because knowledge of the perceptions, beliefs, attitudes, and values of people about important things is highly relevant to policy decisions. We still do not know how to fix the level of a subjective measure quantitatively in immutable units, and I am tempted to say that we can never do that, but I see no point in agonizing over whether the level is correct. In isolation, an absolute measure may be meaningless. But relative differences over time or among groups may be meaningful. I can make little of the information that 43 percent checked one of the boxes labeled "feeling good," "feeling great," or "ecstatic," while 33 percent checked "glum," "despondent," or "despairing." But, if the same cohort of respondents is asked the same question under the same survey conditions a year later and gives quite different answers, I may surmise that something measurable and meaningful has happened.

We are beginning to be able to make comparisons over time and among groups and beginning to be able to conjecture whether the observed differences are real or artifactual. It is harder to identify real changes in subjective than in objective measures. Many, although not all, objective indicators are innately fairly stable. Their measuring devices more likely have been tested and refined over an extended period of time; they more likely have come from larger samples with less sampling variability or from records; and they are more likely resistant to variations in question wording and placement, interviewer bias, and other factors that introduce noise into the measuring system. So, it is not as easy to speak with confidence about changes in trust in government, fear of crime, or intention to vote for a candidate as about changes in the percent of the population over 65, the number of graduate degrees granted, the divorce rate, or even median income.

A study on subjective indicators is underway in the National Research Council under the sponsorship of the National Science Foundation. The study is being conducted under the Committee on National Statistics by its Panel on Survey-Based Measures of Subjective Phenomena, with Otis Dudley Duncan as chairman and Charles Turner as study director. The panel is investigating: (1) the meaning of subjective indicators—what they predict or explain; (2) the reliability and other

statistical characteristics of their measurement (e.g., their error structure and short term stability); (3) the extent to which producers and consumers of such data appreciate their advantages and limitations; and (4) the appropriateness of current practices. The panel's report will assess the nature, extent, and severity of problems affecting subjective survey measurements; suggest ways to deal with some of the problems; and provide guidance for the presentation, interpretation, and use of subjective indicators.

Much of what researchers have accomplished in subjective measurement has been experimental, developmental, and heuristic. We are trying to fix on moving objects with shaky instruments. Still, we are establishing baselines, have achieved some standardization of question wording, and are building a body of information and experience that sets the stage for better detection of trends and variations. Some uncertainties will remain, but practices will become sounder. We will know more about our society.

PENN STATE UNIVERSITY LIBRARIES
A0000709611126

Superintendent of Documents
U.S. Government Printing Office
Washington, D.C. 20402

Official Business
Penalty for Private Use, $300

COM-202
Special Fourth-Class
Rate—Book